

Ein semiparametrisches Verfahren zur
Planung und Auswertung von
Nichtunterlegenheitsstudien im Cox-Modell

DISSERTATION

zur Erlangung des mathematisch-naturwissenschaftlichen
Doktorgrades

“Doctor rerum naturalium”

der Georg-August-Universität Göttingen

vorgelegt von
Karola Kombrink
aus München

Göttingen 2011

D7

Referent: Prof. Dr. Axel Munk

Koreferent: Prof. Dr. Tim Friede

Tag der mündlichen Prüfung: 10.11.2011

Inhaltsverzeichnis

1	Einleitung	1
1.1	Motivation	1
1.2	Inhaltsangabe	3
1.3	Danksagung	6
2	Modell	7
2.1	Einführung in die Überlebenszeitanalyse	7
2.2	Überlebenszeitanalyse bei Weibull-verteilten Überlebenszeiten	16
2.3	Zählprozesse in der Überlebenszeitanalyse	22
3	Schätzer für die Modell-Parameter und ihre asymptotischen Eigenschaften	25
3.1	Schätzung der Modell-Parameter	25
3.2	Asymptotische Eigenschaften des Schätzers	26
3.3	Schätzer der Kovarianzmatrix	31
4	Semiparametrisches Testverfahren	39
4.1	Konzept der Überlegenheit (Vortest)	40
4.2	Konzept der Nichtunterlegenheit	41
4.3	Auswertung der Depressionsstudie	49
5	Konzept der Fallzahlplanung	51
5.1	Berechnung der insgesamt erforderlichen Ereignisse	52
5.2	Berechnung der insgesamt erforderlichen Personen	53

6	Algorithmen zur Bestimmung der Fallzahl	55
6.1	Berechnung der Lage der Ereigniszeitpunkte und ihre insgesamt erforderliche Anzahl	56
6.2	Algorithmen	62
6.2.1	Berechnung der Fallzahlen bei festen Randomisierungsverhältnissen in die Gruppen	62
6.2.2	Berechnung der Fallzahlen mit optimaler Aufteilung in die Gruppen	69
6.3	Untersuchung der Testverfahren mittels Simulationen	75
7	Vergleich mit einem parametrischen Verfahren	81
7.1	Parametrisches Verfahren	82
7.2	Gegenüberstellung des parametrischen und semiparametrischen Verfahrens	84
7.2.1	Exponentialverteilte Überlebenszeiten	84
7.2.2	Weibull-verteilte Überlebenszeiten (Form-Parameter $\vartheta \neq 1$)	89
7.2.3	Auswertung der Depressionsstudie	92
8	Erweiterungen des Modells	97
8.1	Variable Nachbeobachtungszeit	97
8.2	Kovariablen	99
9	Zusammenfassung und Diskussion	101
A	Anhang	103
A.1	103
A.2	106
A.3	110
	Literaturverzeichnis	111

1 Einleitung

In dieser Arbeit entwickeln wir im Rahmen des Cox-Proportional-Hazards-Modells ein semiparametrisches statistisches Verfahren, um eine 3-armige Nichtunterlegenheitsstudie im Bereich der Überlebenszeitanalyse zu planen und auszuwerten. Konkret geht es um die Einführung eines neuen Medikamentes, wenn bereits ein Standardmedikament auf dem Markt existiert gegen das getestet und die Wirksamkeit beider Präparate im direkten Vergleich mit einem Placebo gezeigt werden soll.

1.1 Motivation

Selbst wenn ein etabliertes Standardmedikament (S) existiert, ist die Entwicklung von Alternativpräparaten wünschenswert. So besitzen manche Medikamente starke Nebenwirkungen. Hier wäre ein neues Medikament (N , Novum) mit weniger Nebenwirkungen wünschenswert. Einige Patienten reagieren allergisch auf gewisse Inhaltsstoffe oder sind resistent dagegen. Auch hier ist ein Alternativpräparat unverzichtbar. Ein anderer Vorteil einer neuen Therapie kann darin bestehen, dass sie kostengünstiger als die existierende ist. Für neue Wirkstoffe, die einen der oben genannten Vorteile besitzen, erwartet man in der Regel bezogen auf die eigentliche Krankheit nur geringe Behandlungsunterschiede gegenüber der bereits existierenden Therapie. Würde in dieser Situation eine Überlegenheitsstudie durchgeführt werden, so wären immens hohe Fallzahlen erforderlich, um die Überlegenheit des neuen Medikamentes gegenüber Standard zeigen zu können. Dies führt zu sogenannten mega trials, in denen zehntausende an Personen rekrutiert werden müssen. Die Durchführung solch einer Studie verursacht zum einen extrem hohe Kosten und zum anderen dauern die Studien über mehrere Jahre. Eine Alternative stellt die Durchführung einer Nichtunterle-

genheitsstudie dar. In diesem Fall wird das neue Medikament als vollwertige Alternative zu dem Standardmedikament angesehen, selbst wenn es dem Standardmedikament bezogen auf die eigentliche Krankheit geringfügig unterlegen ist. Die erlaubte Nichtunterlegenheit wird durch eine Nichtunterlegenheitsgrenze quantifiziert, die den Unterschied zwischen Novum und Standard angibt, der gerade noch als klinisch tolerabel angesehen werden kann. Über die Wirksamkeit der beiden Therapien sind allerdings nur Aussagen möglich, wenn zusätzlich zu Standard und Novum ein Placebo (P) mit getestet, also eine 3-armige Studie durchgeführt wird. Solch eine Studie heißt eine Studie im Gold Standard Design (vgl. Koch and Röhmel [2004]). Aus ethischen Gründen ist ein 3-armiges Studiendesign allerdings nicht bei allen Krankheiten geeignet, insbesondere wenn die Schwere der Krankheit den Einsatz von Placebo ausschließt. Jedoch wird es für einige Indikationen wie z.B. die Behandlung von Depression besonders empfohlen (vgl. Röhmel et al. [2005]). Um zu demonstrieren, dass die Studie die Fähigkeit besitzt zwischen wirksamen und unwirksamen Therapien zu unterscheiden (diese Eigenschaft wird assay sensitivity genannt, vgl. beispielsweise Snapinn [2004]), muss im Rahmen der Auswertung der klinischen Studie vorab getestet werden, ob entweder Standard oder Novum besser als Placebo wirkt. Wenn diese Überlegenheit allerdings nicht gezeigt werden kann, darf anschließend nicht mehr auf Nichtunterlegenheit getestet werden, bzw. ist das Testergebnis irrelevant und das neue Medikament dürfte nicht auf dem Markt eingeführt werden.

Sowohl aus ethischer Sicht (ethische Grundsätze für die medizinische Forschung an Menschen sind durch die Declaration of Helsinki, WMA [2008] festgelegt) und Kostengründen als auch aus statistischer Sicht ist die Anzahl der an der Studie teilnehmenden Patienten und deren optimale Aufteilung auf die Behandlungsgruppen besonders sorgfältig zu planen. Einerseits sollen aus den ersten beiden Gründen die Fallzahlen möglichst gering gehalten werden, andererseits sind sie aus dem letzten Grund hinreichend groß festzulegen. Diese Berechnung erfolgt über die asymptotische Güte des zugrundeliegenden Nichtunterlegenheitstests unter Berücksichtigung von möglichen zufällig und/oder fest rechtszensierten Beobachtungen (vgl. Collett [1996], Kapitel 9).

Parametrische Ansätze sind hierzu bereits in der Literatur zu finden. So können für Studien unter der Annahme von exponential- (Mielke et al. [2008a]), normal- (Pigeot et al. [2003], Hasler et al. [2008]), binomial- (Kieser and Friede [2007], Mielke and Munk) oder Poissonverteilten (Mielke and Munk) Überlebenszeiten die erforderlichen Stichprobenumfänge und deren optimale Aufteilung in die Behandlungsgruppen geplant werden. Kann aber z.B. aufgrund der Datenlage nicht angenommen werden, dass die unterstellte Verteilungsannahme

gerechtfertigt ist, dürfen diese Methoden zur Auswertung der Studie nicht herangezogen werden. Aber auch wenn nach der Datenlage eine gewisse Verteilung wahrscheinlich ist, ist letztlich nicht bekannt, welche Verteilung tatsächlich zugrunde liegt.

Dieses motiviert uns einen allgemeineren semiparametrischen Ansatz zu entwickeln. Dabei stellen wir einen Test im Rahmen des Cox-Proportional-Hazards-Modells (nach [Cox \[1972\]](#)) vor, der lediglich zueinander proportionale Hazardfunktionen zur Annahme hat und keine Verteilungsannahme erfordert. Eine Hazardfunktion gibt das Risiko für ein Ereignis zu einem Zeitpunkt t im Verlauf der Studie an, gegeben, dass die Person bis kurz vor t noch an der Studie teilnimmt und das Ereignis bis dahin noch nicht eingetreten ist. Die Bestimmung der erforderlichen Fallzahlen hängt unmittelbar von der erwarteten Lage der Ereigniszeitpunkte ab. Hier sind im Vorfeld entsprechende Annahmen über die Lage der Ereigniszeitpunkte zu treffen. Einerseits besteht der Anspruch, eine möglichst flexible Verteilung zu wählen, mit der wachsende, konstante und auch fallende Hazardfunktionen dargestellt werden können, andererseits ist die Proportionalität der Hazardfunktionen zu gewährleisten. Die Weibull-Verteilung erfüllt diese Eigenschaften. Insbesondere ist die Exponentialverteilung ein Spezialfall der Weibull-Verteilung (Form-Parameter 1), die lediglich konstante Hazardfunktionen abdeckt.

Zur Veranschaulichung der Methoden greifen wir immer wieder auf einen Datensatz einer Studie der Firma Lilly Deutschland GmbH zurück, in der der Verlauf einer Depression bei depressiven Patienten untersucht wurde. Dort wurde ein Medikament mit einem am Markt befindlichen Standardmedikament und einem Placebo in einer doppelblinden, randomisierten Studie verglichen.

1.2 Inhaltsangabe

Die Arbeit ist in 9 Kapitel aufgeteilt. In Kapitel 2 geben wir die wesentlichen Grundlagen aus der Überlebenszeitanalyse an. Mittels Hazardfunktionen im Cox-Proportional-Hazards-Modell untersuchen wir den Einfluss zwischen den Behandlungen und der Zeit, die vergeht, bis das Zielereignis eintritt. Die Hazardfunktionen für jede Behandlungsgruppe sind gegeben durch

$$\begin{aligned} h_P(t) & \quad \text{für Patienten aus der Placebogruppe,} \\ h_S(t) & := h_P(t) \cdot e^{\beta_S} \quad \text{für Patienten aus der Standardgruppe,} \\ h_N(t) & := h_P(t) \cdot e^{\beta_N} \quad \text{für Patienten aus der Novumgruppe,} \end{aligned}$$

wobei β_S, β_N die unbekannt zu schätzenden Regressionskoeffizienten sind, die den Einfluss der Behandlungsgruppen auf die Ereigniszeit repräsentieren.

Kapitel 3 behandelt, wie die Modellparameter β_S, β_N aus den Daten geschätzt werden, ohne dass die Form der Baseline-Hazardfunktion h_P bekannt ist. Da keine spezielle Form für h_P gewählt wird, kann nicht die Maximum-Likelihood-Methode zur Schätzung von β_S und β_N verwendet werden. Als Schätzmethode wird deshalb die Partial-Maximum-Likelihood-Methode verwendet, die sich allerdings aus der Maximum-Likelihood-Methode ableiten lässt. Wir werden sehen, dass der Partial-Maximum-Likelihood-Schätzer in unserem Modell konsistent und unter geeigneten Regularitätsbedingungen asymptotisch normalverteilt ist. Insbesondere geben wir die asymptotische Kovarianzmatrix des Partial-Maximum-Likelihood-Schätzers an. Die einzelnen Elemente dieser Matrix können jedoch nicht explizit berechnet werden, so dass wir für hinreichend große Stichprobenumfänge hierfür eine Approximation angeben. Diese Approximation spielt im weiteren Verlauf der Arbeit eine zentrale Rolle, da sie eine Grundlage für die Fallzahlplanung darstellt. Bevor wir uns mit der Fallzahlplanung beschäftigen, muss zunächst ein Testverfahren entwickelt werden.

In Kapitel 4 stellen wir ein semiparametrisches Testverfahren vor, das aus einem Überlegenheits- und einem Nichtunterlegenheitstest besteht und auf der Partial-Maximum-Likelihood-Methode beruht. Das erste Testverfahren ist bereits aus der Literatur bekannt (vgl. beispielsweise [Bernardo and Harrington \[2001\]](#)) und dient dazu, die assay sensitivity der Studie zu demonstrieren. Für das zweite Testproblem entwickeln wir eine Teststatistik, die auf der Partial-Maximum-Likelihood-Methode beruht und zeigen, dass sie asymptotisch normalverteilt ist. Außerdem untersuchen wir durch Monte-Carlo-Simulationen, ab welchem endlichen Stichprobenumfang bereits eine Normalverteilung angenommen werden kann. Insgesamt betrachten wir dann die folgende Hypothese

$$H_0 : H_0^{Ueb,j} \cup H_0^{Nu} = \{\beta_j \leq 0\} \cup \{\beta_N \leq \Delta \cdot \beta_S\}$$

mit wahlweise entweder $j = S$ oder $j = N$; welche Hypothese dabei konkret gewählt wird, ist a priori festzulegen. In der Nichtunterlegenheitsgrenze $\Delta \in (0, 1)$ ist die klinisch irrelevante Unterlegenheit zwischen N und S quantifiziert. Aufgrund der hierarchischen Struktur dieses Testproblems wird H_0 zum Niveau α abgelehnt, falls sowohl $H_0^{Ueb,j}$ als auch H_0^{Nu} zum Niveau α abgelehnt werden können.

In Kapitel 5 und 6 behandeln wir die Fallzahlplanung. Zunächst stellen wir in Kapitel 5 das allgemeine Vorgehen vor. Die Fallzahlplanung geschieht in dem Sinn, dass bei vorgegebenen Parameterwahlen für das Niveau, die Nichtunterlegenheitsgrenze, Annahmen über die Wirkung der Medikamente zueinander und Annahmen über die Anteile an zensierten

Beobachtungen, der Stichprobenumfang gerade so gewählt wird, dass unter der Alternative $\beta_S = \beta_N \neq 0$ eine gewünschte Güte des Nichtunterlegenheitstests erreicht wird.

In Kapitel 6 wird schließlich ein Algorithmus entwickelt, der es ermöglicht, für unser Modell mit Berücksichtigung von möglichen zensierten Beobachtungen und unter Einbeziehung der Ergebnisse der vorangegangenen Kapitel, geeignete Fallzahlen insgesamt und deren optimale Aufteilung in die Behandlungsgruppen numerisch zu bestimmen. Simulationen bestätigen, dass unter dieser Alternative die simulierte Güte des Tests für H_0 nur unwesentlich kleiner ist als die für H_0^{Nu} . Somit reicht es aus, die Fallzahlplanung basierend auf dem Nichtunterlegenheitstest durchzuführen. Außerdem untersuchen wir die Robustheit des Nichtunterlegenheitstests in Bezug auf das Einhalten des Niveaus und das Erreichen der Güte, wenn die Annahme der proportionalen Hazardfunktionen verletzt ist.

Wie sich unser Modell zu einem bestehenden parametrischen Modell verhält, wird in Kapitel 7 untersucht. Dazu wird das Verfahren von [Mielke et al. \[2008a\]](#), das für exponentialverteilte Überlebenszeiten entwickelt wurde, also auch von proportionalen Hazardfunktionen ausgeht, gewählt. Zunächst vergleichen wir unter der Annahme von exponentialverteilten Daten die Fallzahlen beider Methoden und werden sehen, dass die ermittelten Zahlen kaum voneinander abweichen. Theoretische Berechnungen der relativen asymptotischen Effizienz des semiparametrischen bezüglich des parametrischen Schätzers haben gezeigt, dass diese nahe 1 ist. Anschließend untersuchen wir, wie sich das parametrische Verfahren verhält, wenn die Daten Weibull-verteilt sind mit Form-Parameter $\vartheta \neq 1$. Dort werden wir feststellen, dass das parametrische Verfahren nicht robust bezüglich der Verletzung der Verteilungsannahme ist, während unsere Methode weiterhin das Niveau hält und die gewünschte Güte gut erzielt und das sogar, obwohl wir die Fallzahlen von [Mielke et al. \[2008a\]](#) zugrunde gelegt haben.

In Kapitel 8 geben wir Hinweise auf möglichen Modellerweiterungen. Konkret betrachten wir variable Nachbeobachtungszeiten der Patienten und berücksichtigen zusätzliche Kovariablen.

Abschließend folgt in Kapitel 9 eine Zusammenfassung und Diskussion der Ergebnisse dieser Arbeit.

1.3 Danksagung

An dieser Stelle möchte ich Herrn Prof. Dr. Axel Munk für die Vergabe des interessanten Forschungsthemas und seine vielseitige Unterstützung ganz herzlich danken. Auch danke ich Herrn Prof. Dr. Tim Friede von der Abteilung Medizinische Statistik für seine wertvollen Anregungen, vor allem auch zum praktischen und medizinischen Hintergrund des Themas. Dem Vorstand des Zentrums für Statistik und des Graduiertenkollegs 1023 danke ich für die Gewährung eines Stipendiums und die Aufnahme in die jeweiligen Promotionsprogramme, welche mir interessante Einblicke in andere Themengebiete ermöglicht haben. Ebenfalls zu Dank verpflichtet bin ich der Firma Lilly Deutschland GmbH für die Bereitstellung des Datensatzes.

Für die stets gute Arbeitsatmosphäre am Institut für Mathematische Stochastik möchte ich mich bei allen Institutskollegen und insbesondere bei meinen Bürokollegen bedanken. Ein besonderer Dank gilt meinen Eltern und meiner Schwester Sabrina, die mir stets zur Seite gestanden haben.

2 Modell

2.1 Einführung in die Überlebenszeitanalyse

Um die Wirkung eines neuen Medikamentes zu testen, soll dieses mit der Wirkung eines bereits am Markt befindlichen Standardmedikamentes sowie eines Placebos im Rahmen einer klinischen Studie verglichen werden. Vor der Durchführung der Studie sind die Dauer der Studie und die Anzahl der an ihr teilnehmenden Patienten pro Gruppe festzulegen. Weil die Personenanzahl großen Einfluss auf die Aussagekraft der Studie und die mit der Studie verbundenen Kosten besitzt, ist ihre sorgfältige Planung von großer Bedeutung. Doch nicht nur aus Kostengründen, sondern auch aus ethischer Sicht ist ein unnötig großer Stichprobenumfang zu vermeiden. Werden hingegen zu wenig Patienten betrachtet, so kann es sein, dass zu wenig Informationen in den Daten enthalten sind, um ein signifikantes Ergebnis zu erhalten. Dadurch kann es passieren, dass das Studienziel nicht erreicht wird. Aus diesem Interessenkonflikt ergibt sich in Bezug auf die Fallzahlen also die Fragestellung: Wie viele Patienten müssen in welcher Behandlungsgruppe insgesamt beobachtet werden, um mit einer bestimmten Wahrscheinlichkeit ein signifikantes Ergebnis nachzuweisen?

\mathbb{B} : sei der *Beobachtungszeitraum*, in dem die Personen in der Studie beobachtet werden. Die maximale Verweildauer jedes Patienten in der Studie beträgt $b \in \mathbb{R}_{>0}$ Zeiteinheiten, also gilt $\mathbb{B} := [0, b]$.

n : sei die *Anzahl der Personen*, die insgesamt an der Studie teilnehmen. Die Personen seien voneinander unabhängig.

Ebenfalls vor der Durchführung der Studie teilen wir die Patienten in drei Gruppen ein, die sich durch die jeweils verabreichte Medikation unterscheiden: Personen aus Gruppe 1 bekommen das Standardmedikament (S), aus Gruppe 2 das Novum (N) - die neu entwickelte Therapie - und aus Gruppe 3 das Placebo (P). Die Gruppengrößen bezeichnen wir wie folgt:

n_k : sei die Anzahl der Personen in Gruppe k , wobei $k = S, N, P$.

Es gilt also $n = n_S + n_N + n_P$, wobei die n_k verschieden sein können.

$w_k = n_k/n \in (0, 1)$ bezeichne den Anteil der Personen aus Gruppe $k = S, N, P$ an der gesamten Personenanzahl.

Als Kriterium zur Beurteilung der Wirksamkeit einer Medikation werten wir Zeiten bis zum Auftreten eines Zielereignisses aus. Die dafür notwendige Notation stellen wir nun vor. Im Folgenden sei stets $i \in \{1, \dots, n\}$, $k = S, N, P$ und $t \in \mathbb{B}$.

T_i : sei eine stetige, nichtnegative Zufallsvariable, die die Überlebenszeit von Person i beschreibt. Für diese Zufallsvariable fordern wir $P(T_i \geq b) > 0$. Außerdem seien die Zufallsvariablen T_1, \dots, T_n unabhängig.

L : sei die Anzahl der Personen, bei denen im Verlauf der Studie das Zielereignis eintritt. Es gilt $L \leq n$.

$T_{(l)}$: sei die l -te Ordnungsstatistik der Stichprobe $\{T_i\}$. Wir nehmen an, dass pro Zeitpunkt stets nur ein Ereignis eintritt. In diesem Fall sagt man, dass keine Bindungen vorliegen.

t_l : sei für $l \in \{1, \dots, L\}$ die Realisierung der Zufallsvariable $T_{(l)}$.

F_k : sei die Verteilungsfunktion der Überlebenszeit einer beliebigen Person aus Gruppe k . Wir nehmen an, dass die Verteilung der Überlebenszeit innerhalb jeder Behandlungsgruppe gleich ist.

f_k : sei die Dichtefunktion der Überlebenszeit einer beliebigen Person aus Gruppe k . Es gelte $f_k(t) > 0$ für alle $t \in \mathbb{B}$.

Da in der Regel während der Studie nicht bei allen Patienten das Zielereignis beobachtet wird, ist die exakte Überlebenszeit oft unbekannt. Man spricht in diesem Fall von einer zensierten Überlebenszeit (vgl. z.B. Klein and Moeschberger [2005], Kapitel 3). Speziell spricht man von einer zufälligen Rechtszensierung, wenn die Beobachtung eines Patienten

innerhalb des Beobachtungszeitraumes aus einem Grund abgebrochen wird, den der Beobachtende nicht beeinflussen kann. Gründe dafür sind zum Beispiel, dass der Patient in eine andere Stadt zieht, die weitere Teilnahme verweigert oder durch einen Unfall verstirbt, wobei der Grund des Ausscheidens nicht mit der Therapie in Zusammenhang zu bringen ist. Ein anderer Spezialfall der Rechtszensierung ist die *festе Rechtszensierung*: Da der Beobachtungszeitraum in der Praxis begrenzt ist, kann es sein, dass das Zielereignis einer Person erst nach dem Ende des Beobachtungszeitraumes eintritt. Für diese Personen ist der Endzeitpunkt das Studienende. Da die beobachteten zensierten Überlebenszeiten ebenfalls Informationen beisteuern, sollen sie auch in unserem Modell mit berücksichtigt werden.

U_i : sei eine stetige, nichtnegative Zufallsvariable, die die *zufällige Rechtszensierungszeit* von Person i beschreibt. Es gelte $P(U_i \geq b) > 0$. Die Zufallsvariablen U_1, \dots, U_n seien unabhängig. Für festes i seien T_i und U_i unabhängig.

F_{U_1} : sei die *Verteilungsfunktion der zufälligen Rechtszensierungszeit*. Wir nehmen an, dass die Verteilung dieser Zensierungszeiten aller Personen gleich ist.

Die Wahrscheinlichkeit für eine fest bzw. zufällig rechtszensierte Beobachtung sei für jede Gruppe folgendermaßen bezeichnet:

$p_f^k := P(b < T_k, b < U_1) \in (0, 1)$ sei die *Wahrscheinlichkeit* für eine *fest rechtszensierte* Beobachtung einer Person aus Gruppe k .

$p_z^k := P(U_1 < T_k, U_1 < b) \in (0, 1)$ sei die *Wahrscheinlichkeit* für eine *zufällig rechtszensierte* Beobachtung einer Person aus Gruppe k .

$\tilde{U}_i := \min\{U_i, b\}$ sei die *Zensierungszeit* von Person i .

$\delta_i := \mathbb{1}_{\{T_i < \tilde{U}_i\}} = \begin{cases} 1 & \text{bei Person } i \text{ tritt während der Studie das Zielereignis ein} \\ 0 & \text{Person } i \text{ ist zensiert} \end{cases}$

bezeichnet den *Zensierungsindikator* von Person i .

Die beobachteten Daten bestehen für jede Person i aus dem Paar $(\min\{T_i, \tilde{U}_i\}, \delta_i)$.

Nachdem wir nun die grundlegende Notation eingeführt haben, führen wir zwei wichtige Größen aus der Überlebenszeitanalyse ein, mit Hilfe derer Aussagen über die Wirkung der Therapien getroffen werden können. Weitere Grundlagen aus der Überlebenszeitanalyse sind beispielsweise in [Therneau and Grambsch \[2001\]](#) zu finden.

Um den Einfluss von Behandlungsmethoden auf die Zeit, die vergeht, bis das Zielereignis eintritt zu untersuchen, betrachten wir das Risiko für das Eintreten eines Ereignisses zum Zeitpunkt t , gegeben, dass die Person bis kurz vor t noch an der Studie teilnimmt und das Ereignis bis dahin noch nicht eingetreten ist. Dies kann mittels der Hazardfunktion ausgedrückt werden.

Definition 2.1 (Hazardfunktion, Hazard Quotient)

Die Funktion, die für jeden Zeitpunkt t des Beobachtungszeitraumes für den beobachteten Patienten i das momentane Ereignisrisiko angibt, gegeben, dass die Person bis zur Zeit t überlebt hat, wird *Hazardfunktion* bezüglich der Überlebenszeit T_i genannt und mit \tilde{h}_i bezeichnet. In Formeln:

$$\tilde{h}_i(t) := \lim_{dt \downarrow 0} \frac{P(T_i \in [t, t + dt) | T_i \geq t)}{dt}.$$

Unterscheiden sich zwei Hazardfunktionen im gesamten Zeitverlauf nur durch einen konstanten Faktor, ist ihr Quotient also konstant, so nennt man sie *proportionale Hazardfunktionen*; ihr konstanter Quotient heißt *Hazard Quotient* ($=:HQ$).

Ein konstanter Hazard Quotient bedeutet, dass das Verhältnis der Therapieeffekte, gemessen mittels Hazardfunktionen, gleichmäßig über die Zeit besteht. In unserem Modell wollen wir hiervon ausgehen und den Einfluss der Behandlungen mittels Hazard Quotienten ausdrücken.

[Rothmann et al. \[2001\]](#) nennen Gründe, die den Ansatz unterstützen Testverfahren basierend auf Hazard Quotienten zu entwickeln:

- „a hazard ratio seems to be less population dependent than the absolute difference“,
- „in the case of oncology, the hazard ratio is the usual way of comparing treatments with respect to time to event (time to progression in survival)“,
- „the proportional hazards assumption is regarded as reasonable (approximately holding)“.

Wir betrachten die Hazardfunktionen im Cox-Proportional-Hazards-Modell (kurz Cox-Modell, vgl. Cox [1972] und Cox [1975]). Dieses Modell berücksichtigt die Annahme, dass der Quotient zweier Hazardfunktionen über die komplette Studiendauer konstant ist und auch Informationen aus zensierten Beobachtungen nicht verloren gehen. Bei dem Cox-Modell handelt es sich um ein Regressionsmodell für zensierte Daten, bei dem angenommen wird, dass die Hazardfunktion bezüglich der Überlebenszeit T_i von Person i mit der Kovariablen Z_i von folgender Form ist:

$$h_i(t) := h(t|Z_i) := \lim_{dt \downarrow 0} \frac{P(T_i \in [t, t + dt] | T_i \geq t, Z_i)}{dt} = h_0(t) \cdot e^{\beta \cdot Z_i}. \quad (2.1)$$

Mit Hilfe des Cox-Modells lässt sich der Einfluss von erklärenden Variablen Z_i auf die Überlebenszeit untersuchen, wobei die auftretenden Variablen wie folgt definiert sind:

$Z_i := (Z_{iS}, Z_{iN})^T$ sei der *Kovariablenvektor*, der die Information welcher Therapiegruppe der jeweilige Patient i angehört, beinhaltet:

$$Z_{iS} := \begin{cases} 1 : \text{Person } i \text{ nimmt } S \\ 0 : \text{sonst} \end{cases} \quad Z_{iN} := \begin{cases} 1 : \text{Person } i \text{ nimmt } N \\ 0 : \text{sonst.} \end{cases} \quad (2.2)$$

Somit lautet der Kovariablenvektor

$$Z_i = \begin{cases} (1, 0)^T : & \text{falls Person } i \text{ aus der Standardgruppe stammt,} \\ (0, 1)^T : & \text{falls Person } i \text{ aus der Novumgruppe stammt und,} \\ (0, 0)^T : & \text{falls Person } i \text{ aus der Placebogruppe stammt.} \end{cases}$$

Der Wert des Kovariablenvektors ist also für jeden Patienten während der gesamten Studiendauer konstant.

$\beta := (\beta_S, \beta_N) \in \mathbb{R}^2$ sei der unbekannte und zu schätzende *Vektor der Regressionskoeffizienten*, der den Einfluss des Kovariablenvektors Z_i auf die Ereigniszeit repräsentiert.

h_0 : bezeichne die Hazardfunktion der Personen aus der Referenzgruppe. Sie wird *Baseline-Hazardfunktion* genannt. In unserem Fall entspricht die Baseline-Hazardfunktion der Hazardfunktion unter Placebo und wird mit h_P bezeichnet. Da wir den Quotienten von Hazardfunktionen betrachten, können wir somit (wie wir später sehen werden) h_0 unspezifiziert lassen. In diesem Fall sprechen wir von einem semiparametrischen Ansatz.

Im Cox-Modell haben unsere Hazardfunktionen für jede Person i die folgende Form:

$$h_i(t) = h_P(t) \cdot e^{\beta_S \cdot Z_{iS} + \beta_N \cdot Z_{iN}}. \quad (2.3)$$

Die Hazardfunktionen für jede Behandlungsgruppe sind speziell für unser Setting gegeben durch

$$\begin{aligned} h_P(t) & \quad \text{für Patienten aus der Placebogruppe,} \\ h_S(t) & := h_P(t) \cdot e^{\beta_S} \quad \text{für Patienten aus der Standardgruppe und} \\ h_N(t) & := h_P(t) \cdot e^{\beta_N} \quad \text{für Patienten aus der Novumgruppe.} \end{aligned}$$

Im Regressionsvektor β ist die Information über das Therapieverhältnis enthalten, denn

$$\begin{aligned} \beta_S & = \log \left(\frac{h_S(t)}{h_P(t)} \right) \Leftrightarrow HQ(S, P) := \frac{h_S(t)}{h_P(t)} = e^{\beta_S}, \\ \beta_N & = \log \left(\frac{h_N(t)}{h_P(t)} \right) \Leftrightarrow HQ(N, P) := \frac{h_N(t)}{h_P(t)} = e^{\beta_N}. \end{aligned} \quad (2.4)$$

Der Wert von β_N kann folgendermaßen interpretiert werden (analoges gilt für β_S):

- Ist $\beta_N = 0$, also $\frac{h_N(t)}{h_P(t)} = 1$, so unterscheiden sich das neue Medikament und Placebo nicht in ihrer Wirkung in Hinblick auf das Zielereignis.
- Ist $\beta_N < 0$, also $\frac{h_N(t)}{h_P(t)} < 1$, so ist das momentane Ereignisrisiko unter Novum kleiner als unter Placebo.
- Ist $\beta_N > 0$, also $\frac{h_N(t)}{h_P(t)} > 1$, so ist das momentane Ereignisrisiko unter Novum größer als unter Placebo.

Ist das Ereignis z.B. der Tod, so soll die Zeit bis zum Eintritt des Ereignisses möglichst lang sein, d.h. je kleiner die Werte der Hazardfunktion sind, desto besser ist die Therapie. Hier wäre also $\beta_N < 0$ wünschenswert.

Ist das Ereignis z.B. das Nachlassen eines Schmerzes oder der Rückgang einer Depression, so soll die Zeit bis zum Eintritt dieses Ereignisses möglichst kurz sein, d.h. je größer die Werte der Hazardfunktion sind, desto besser ist die Therapie. Hier wäre also $\beta_N > 0$ wünschenswert.

Die folgende Funktion stellt eine wichtige Größe in der Überlebenszeitanalyse dar, die in direkter Beziehung zur Hazardfunktion steht.

Bemerkung 2.2 (Überlebenszeitfunktion)

Die Funktion $S_i(t) := S(t|Z_i) := 1 - F_i(t) = P(T_i \geq t|Z_i) \in [0, 1]$ heißt *Überlebenszeitfunktion* von Person i . Sie gibt die Wahrscheinlichkeit an, bis t oder länger als t zu überleben. S_i ist eine monoton fallende Funktion mit $S_i(0) = 1$ und $\lim_{x \rightarrow \infty} S_i(x) = 0$.

Es gilt folgende Beziehung:

$$S_i(t) = e^{-\int_0^t h_i(u) du} = [S_P(t)]^{e^{\beta Z_i}} = \frac{f_i(t)}{h_i(t)}.$$

Die Überlebenszeitfunktion kann durch die Kaplan-Meier Kurve geschätzt werden. Hierbei handelt es sich um ein nichtparametrisches Verfahren.

Definition 2.3 (Kaplan-Meier Schätzer)

Seien $t_{(1)} < t_{(2)} < \dots < t_{(r)}$ die $r \leq L$ beobachteten unterschiedlichen nach der Größe geordneten Ereigniszeitpunkte. Außerdem sei $t_{(0)}$ der Zeitpunkt des Studienanfangs und $t_{(r+1)} := b$ der Zeitpunkt des Studienendes. Für $j \in \{1, \dots, r\}$ sei mit d_j die Anzahl der Ereignisse zum Zeitpunkt $t_{(j)}$ und mit e_j die Anzahl der Patienten, von denen bekannt ist, dass bei ihnen bis kurz vor dem Zeitpunkt $t_{(j)}$ das Zielereignis noch nicht eingetreten ist, bezeichnet. Der Kaplan-Meier Schätzer der Überlebenszeitfunktion ist folgendermaßen definiert:

$$\hat{S}(t) := \begin{cases} \prod_{j=1}^k \frac{e_j - d_j}{e_j} & : t \in [t_{(k)}, t_{(k+1)}) \text{ mit } 1 \leq k \leq r \\ 1 & : t < t_{(1)}. \end{cases}$$

Depressionsstudie

Uns liegt ein Datensatz einer Studie der Firma Lilly vor, in der der Verlauf einer Depression bei depressiven Patienten untersucht wurde. Typischerweise setzt die Wirkung eines Antidepressivums erst zwei Wochen nach der ersten Einnahme ein. Bis die Wirkung einsetzt, bleiben die Patienten also weiterhin depressiv. Das primäre Ziel der vorliegenden Studie war, den Zeitpunkt der antidepressiven Wirkung von einem Novum mit einem Standardpräparat und einem Placebo innerhalb der ersten beiden Wochen nach Einnahme miteinander zu vergleichen. Sekundäre Ziele waren Aussagen über die Wirkung der Therapien basierend auf dem Fragebogen Hamilton Depression Rating Scale (17 Item Version, kurz HAMD₁₇) während der gesamten 10-wöchigen Studienzeit zu treffen. Um diesen Wert, der ein Maß für die Schwere der Depression darstellt, zu erhalten, beantwortete jeder Patient während der Studie zu 9 verschiedenen Zeitpunkten 17 Fragen (dies geschah die ersten 7 mal im Abstand von einer Woche, dann 2 mal im zwei wöchentlichen Abstand). Falls der HAMD₁₇-Wert unter 7 ist, wird die Person als nicht depressiv eingestuft (hierbei handelt es sich um einen klinisch relevanten Endpunkt, vgl. [Montgomery \[1994\]](#)). Wir wollen für unsere Analyse alle uns zur Verfügung stehenden Daten verwenden und betrachten somit das sekundäre Studienziel.

Wir werden diese Depressionsstudie an dieser und auch späteren Stellen immer wieder zur Veranschaulichung der Methoden heranziehen. Dann haben wir es mit Depressionsstudie (Fortsetzung) bezeichnet.

Insgesamt liegen Daten von 664 Patienten vor. Davon erhielten 267 Patienten Standard, 262 Patienten Novum und 135 Patienten Placebo. Nach den 10 Wochen ist davon bei 311 Patienten das Zielereignis (Remission) eingetreten. Bei den anderen Personen handelt es sich um rechtszensierte Beobachtungen. Der Zensierungsanteil in der Gruppe, die das Standard eingenommen hat, beträgt 54%, in der die Novum nahm 49% und in der Placebogruppe 59%.

Um die vorhandenen Daten graphisch zu veranschaulichen, haben wir in [Abbildung 2.1](#) für jede Behandlungsgruppe die nach der Kaplan-Meier Methode geschätzten Überlebenszeitfunktionen gezeichnet. Die Zensierungszeitpunkte sind durch Kreuze markiert.

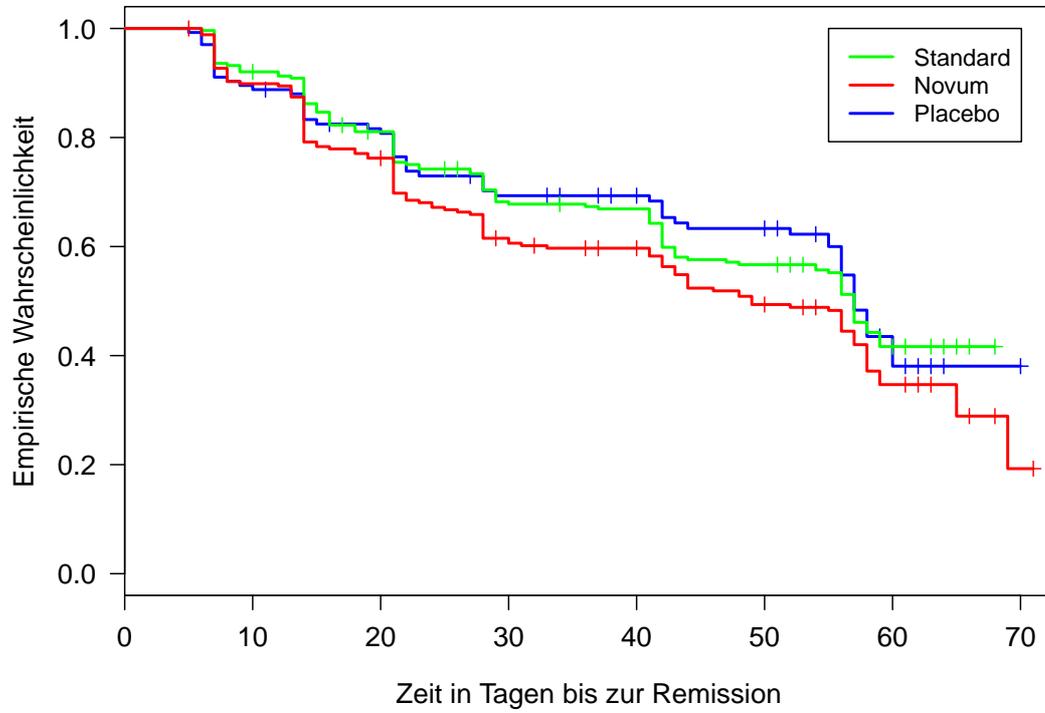


Abbildung 2.1: Nach der Kaplan-Meier Methode geschätzte Überlebenszeitfunktionen der drei Medikamentengruppen aus der Depressionsstudie.

Man sieht, dass die geschätzte Überlebenszeitfunktion der Novumgruppe ab Tag 14 stets unterhalb der geschätzten Überlebenszeitfunktion der Standard- und der Placebogruppe verläuft. Können wir annehmen, dass das neue Medikament eine Alternative zum Standardmedikament darstellt, oder ist die Abweichung rein zufälliger Natur und lässt keine signifikante Aussage zu? Dies wollen wir im weiteren Verlauf der Arbeit untersuchen und hierauf eine Antwort geben.

2.2 Überlebenszeitanalyse bei Weibull-verteiltern Überlebenszeiten

Die Lage der Ereigniszeitpunkte soll durch eine möglichst allgemein gefasste Familie von Verteilungsfunktionen beschrieben werden, deren Hazardfunktionen proportional zueinander sind. Wir zeigen, dass die Familie der Weibull-Verteilungen diese Eigenschaft für uns in geeigneter Weise erfüllt. Weibull-Verteilungen werden durch einen Skalen-Parameter $\lambda \in \mathbb{R}_{>0}$ und einen Form-Parameter $\vartheta \in \mathbb{R}_{>0}$ charakterisiert; z.B. ergibt der Form-Parameter 1 die Exponentialverteilung.

Die Dichte, Verteilungs- und Hazardfunktion von $T_i \sim Wb(\lambda_i, \vartheta_i)$ sind gegeben durch:

$$\begin{aligned} f_i(t) &= \lambda_i \vartheta_i t^{\vartheta_i-1} e^{-\lambda_i t^{\vartheta_i}}, \\ F_i(t) &= 1 - e^{-\lambda_i t^{\vartheta_i}}, \\ h_i(t) &= \lambda_i \vartheta_i t^{\vartheta_i-1}. \end{aligned}$$

Folgen die Überlebenszeiten von Patienten jeder Therapiegruppe einer Weibull-Verteilung mit jeweils gleichem Form-Parameter $\vartheta_i = \vartheta$ für alle i und beliebigem Skalen-Parameter λ_i , so sind die Hazardfunktionen von Personen aus verschiedenen Therapiegruppen proportional zueinander. Betrachten wir nämlich den Quotienten zweier Hazardfunktionen h_i und h_j mit $i, j \in \{1, \dots, n\}$ und $i \neq j$, so sieht man die Proportionalität der Hazardfunktionen leicht:

$$\frac{h_i(t)}{h_j(t)} = \frac{\lambda_i \vartheta t^{\vartheta-1}}{\lambda_j \vartheta t^{\vartheta-1}} = \frac{\lambda_i}{\lambda_j}.$$

Um einen Eindruck über den Formenreichtum von Hazardfunktionen bei Weibull-verteiltern Überlebenszeiten zu bekommen, sind in [Abbildung 2.2](#) Hazardfunktionen für verschiedene Skalen- und Form-Parameter abgebildet.

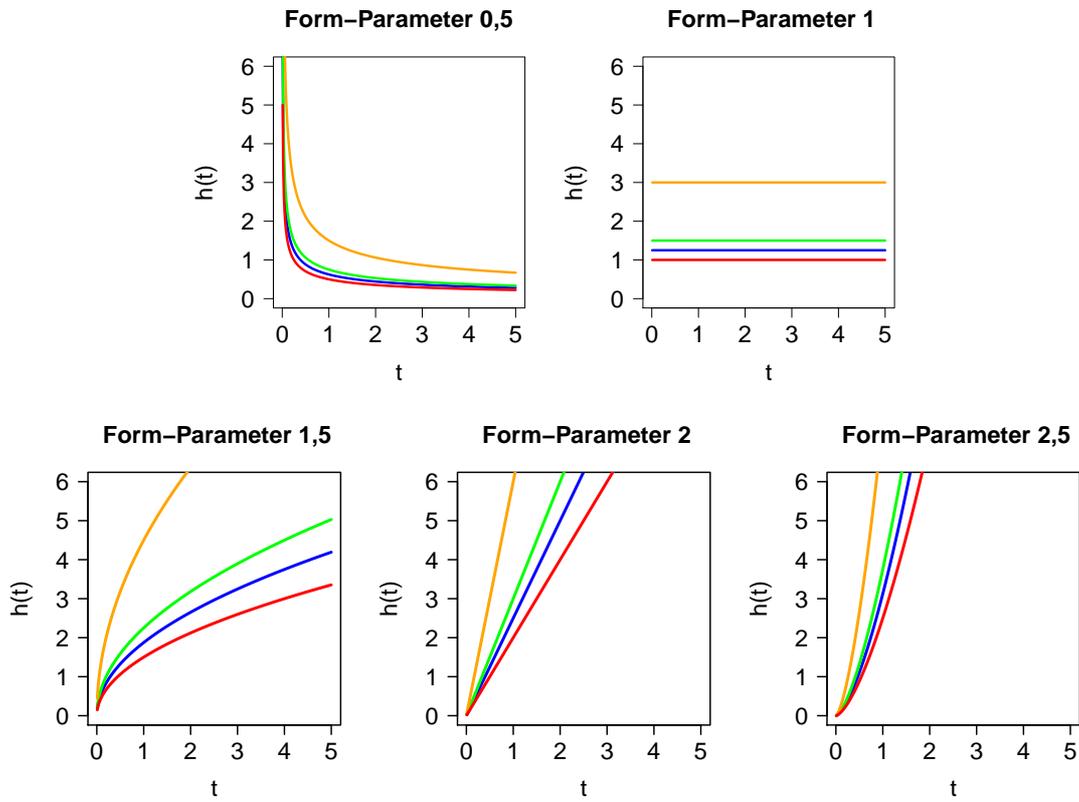


Abbildung 2.2: Hazardfunktionen bei Weibull-verteiltern Überlebenszeiten mit Skalen-Parameter $\lambda \in \{1, 1,25, 1,5, 3\}$ für verschiedene Form-Parameter.

In jeder der fünf Abbildungen sind die Form-Parameter jeweils gemäß Überschrift gleich gewählt, so dass die Hazardfunktionen pro Abbildung proportional zueinander sind. Für einen festen Form-Parameter gilt für festes t folgendes: je kleiner der Skalen-Parameter ist, desto geringer ist das momentane Ereignisrisiko. Wir stellen fest, dass durch geeignete Wahl der Form-Parameter viele Typen von Hazardfunktionen dargestellt werden können, nämlich:

1. *Streng monoton fallende Hazardfunktion* (vgl. Abbildung 2.2 Form-Parameter 0,5)

Wird der Form-Parameter aus dem Intervall $(0, 1)$ gewählt, so fällt das momentane Ereignisrisiko mit zunehmendem t zunächst schnell und dann immer langsamer. Beispielsweise könnte so der Verlauf des Todesrisikos bei Personen nach einer Operation aussehen. Da das Risiko an den Folgen einer Operation zu versterben gewöhnlicherweise geringer wird, je länger die Operation zurückliegt, wird hier mit einer immer schwächer fallenden Ereignisrate gerechnet (nach Kleinbaum [1999], Kapitel 1).

2. Konstante Hazardfunktion (vgl. Abbildung 2.2 Form-Parameter 1)

Beträgt der Form-Parameter 1, so ist das Ereignisrisiko während der Studiendauer konstant. Beispielsweise können hierdurch Risikoverläufe bei Krankheiten modelliert werden, bei denen das Ereignis zu jedem Zeitpunkt mit der gleichen Intensität eintritt.

3. Streng monoton wachsende Hazardfunktion

Hier unterteilen wir in drei Typen:

- Für einen Form-Parameter zwischen 1 und 2 wächst das Ereignisrisiko im Studienverlauf zunächst schnell und dann immer langsamer (vgl. Abbildung 2.2 Form-Parameter 1,5).
- Wählt man 2 als Form-Parameter, so sind die resultierenden Hazardfunktionen linear (vgl. Abbildung 2.2 Form-Parameter 2).
- Ist der Form-Parameter größer als 2, wächst das Ereignisrisiko im Studienverlauf zunächst langsam und dann immer schneller (vgl. Abbildung 2.2 Form-Parameter 2,5).

Soll die Zeit bis zum Eintreten des Todes untersucht werden, so wäre beispielsweise ein wachsendes Ereignisrisiko bei an Krebs erkrankten Patienten zu erwarten, die kaum oder gar nicht auf die Behandlung ansprechen. Hier steigt aufgrund der fortschreitenden Krankheit das Todesfallrisiko im Laufe der Zeit immer weiter an (nach Kleinbaum [1999], Kapitel 1).

Somit können durch die Weibull-Verteilung in der Tat viele mögliche Fallgestaltungen abgedeckt werden, bei stets proportionalen Hazardfunktionen. Im Folgenden werden wir die drei Haupttypen von Hazardfunktionen untersuchen und dazu die Form-Parameter $\vartheta \in \{0,5; 1; 1,5\}$ betrachten.

Depressionsstudie (Fortsetzung)

Für unseren Datensatz wollen wir nun überprüfen, ob die Daten der Gruppen als Weibullverteilt mit einem gemeinsamen Form-Parameter angenommen werden können und somit proportionale Hazardfunktionen vorliegen. Dafür passen wir an die Daten Weibull-Verteilungen mit verschiedenen Form-Parametern pro Gruppe (Modell 1) bzw. mit demselben Form-Parameter für alle Gruppen (Modell 2) an. Die Parameter der Weibull-Verteilungen schätzen wir mit der Maximum-Likelihood-Methode. Schließlich vergleichen wir beide Modelle miteinander.

Modell 1: Zunächst passen wir an die Daten jeder Behandlungsgruppe eine Weibull-Verteilung an und erhalten folgende Schätzer für die Parameter der Weibull-Verteilungen:

$$\begin{aligned}\hat{\lambda}_{S,1} &= 0,00361, & \hat{\vartheta}_{S,1} &= 1,312, \\ \hat{\lambda}_{N,1} &= 0,00515, & \hat{\vartheta}_{N,1} &= 1,279, \\ \hat{\lambda}_{P,1} &= 0,00543, & \hat{\vartheta}_{P,1} &= 1,189.\end{aligned}$$

Modell 2: Jetzt passen wir Weibull-Verteilungen an die Daten an, diesmal aber unter der Annahme, dass die Form-Parameter der Behandlungsgruppen übereinstimmen. Die aus den Daten geschätzten Parameter lauten nun:

$$\hat{\lambda}_{S,2} = 0,00418, \quad \hat{\lambda}_{N,2} = 0,00525, \quad \hat{\lambda}_{P,2} = 0,00391 \quad \text{und} \quad \hat{\vartheta}_2 = 1,2743.$$

Um die Güte der Anpassung an die Daten beurteilen zu können, betrachten wir P-Plots. Dabei gehören die grünen Kreuze zu Modell 1 und die roten Kreise zu Modell 2. In Abbildung 2.3 ist für die Novumgruppe der P-Plot gezeichnet.

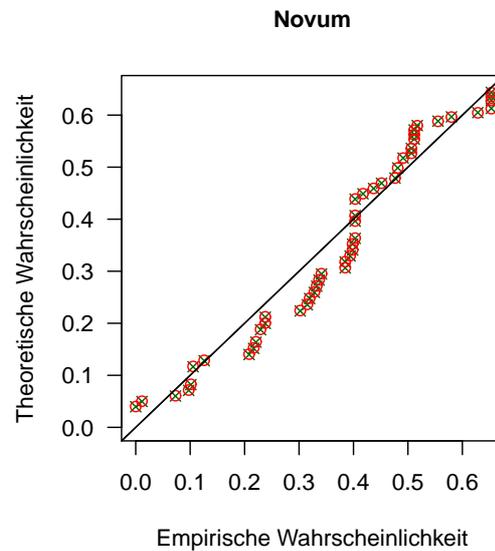


Abbildung 2.3: P-Plot für **Modell 1** grüne Kreuze, **Modell 2** rote Kreise

Die roten Kreise liegen ziemlich genau auf den grünen Kreuzen. Die Anpassungen beider Modelle sind hier somit gleich gut. Insgesamt weichen die Kreuze und Kreise nur schwach von der Ursprungsgeraden mit Steigung 1 ab. Daher können wir die Daten der Novumgruppe jeweils als Weibull-verteilt mit den oben angegebenen Parametern annehmen. In

Abbildung 2.4 sind für die Standard- und Placebogruppe die P-Plots gezeichnet.

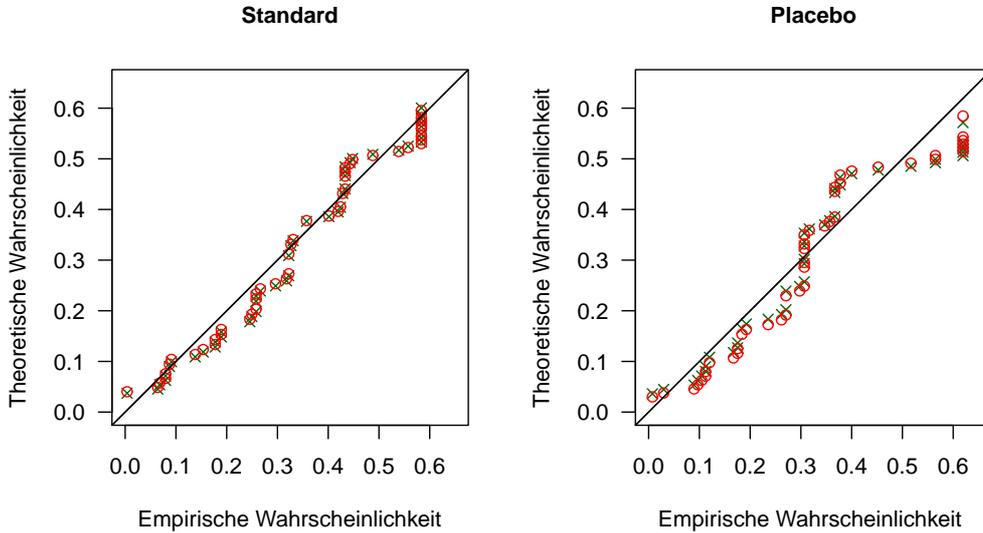


Abbildung 2.4: P-Plots für **Modell 1** grüne Kreuze, **Modell 2** rote Kreise

Kreuze und Kreise liegen auch hier jeweils nah beieinander. Für die Standardgruppe liegen die Kreise etwas näher an der Ursprungsgeraden mit Steigung 1 als die Kreuze. Für die Placebogruppe sind die größten Abweichungen zwischen Kreuzen zu Kreisen zu beobachten. Hier liegen mal die Kreuze und mal die Kreise näher an der Ursprungsgeraden. Allerdings weichen die Kreise und Kreuze jeweils nur schwach von der Ursprungsgeraden mit Steigung 1 ab. Daher können wir die Daten in allen Fällen als Weibull-verteilt mit den oben angegebenen Parametern annehmen. Dies lässt bereits vermuten, dass beide Modelle gleichwertig sind und wird nun noch mit einem Test untersucht.

Vergleich der beiden Modelle mittels Likelihood-Quotienten-Test

Sei $\Theta := \{(\lambda_S, \lambda_N, \lambda_P, \vartheta_S, \vartheta_N, \vartheta_P) : \lambda_S, \lambda_N, \lambda_P, \vartheta_S, \vartheta_N, \vartheta_P \in \mathbb{R}_{>0}\}$ der Parameterraum und $\Theta_0 := \{(\lambda_S, \lambda_N, \lambda_P, \vartheta_S, \vartheta_N, \vartheta_P) : \lambda_S, \lambda_N, \lambda_P, \vartheta_S = \vartheta_N = \vartheta_P \in \mathbb{R}_{>0}\}$ der Hypothesenraum.

Mit \hat{L}_j sei der maximale Wert der Likelihoodfunktion von Modell j bezeichnet, $j = 1, 2$.

Die beiden Modelle können nun im Rahmen des Likelihood-Quotienten-Tests anhand ihrer Werte von \hat{L}_j miteinander verglichen werden. Als Niveau wählen wir 5%.

Um die Hypothese $H_0 : \theta \in \Theta_0$ zu testen, verwenden wir die Teststatistik

$$\tilde{T} := -2 \log \hat{L} = 2 \left(\log \hat{L}_1 - \log \hat{L}_2 \right) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \chi_{\dim \Theta - \dim \Theta_0}^2 = \chi_2^2.$$

Da in Modell 1 mehr Parameter enthalten sind als in Modell 2, ist \hat{L}_1 größer als \hat{L}_2 . Wir wollen nun untersuchen, ob es ausreicht Modell 2 mit der geringeren Anzahl an Parametern zu verwenden. Für den Datensatz ist $\tilde{T} = 0,5$. Der zu diesem Wert der Teststatistik gehörende p-Wert beträgt 77,9%. Daher kann die Hypothese zum 5% Niveau nicht verworfen werden. Insgesamt können wir also die beiden Modelle als gleichwertig ansehen und ziehen das Modell 2 dem Modell 1 vor. Somit ist die Annahme gerechtfertigt, dass die an die Daten angepassten Weibull-Verteilungen denselben Form-Parameter besitzen und die resultierenden Hazardfunktionen proportional zueinander sind.

In Abbildung 2.5 sind die Kaplan-Meier Kurven zusammen mit Überlebenszeitkurven von Weibull-Verteilungen mit den oben angegebenen geschätzten Parametern aus Modell 2 eingetragen.

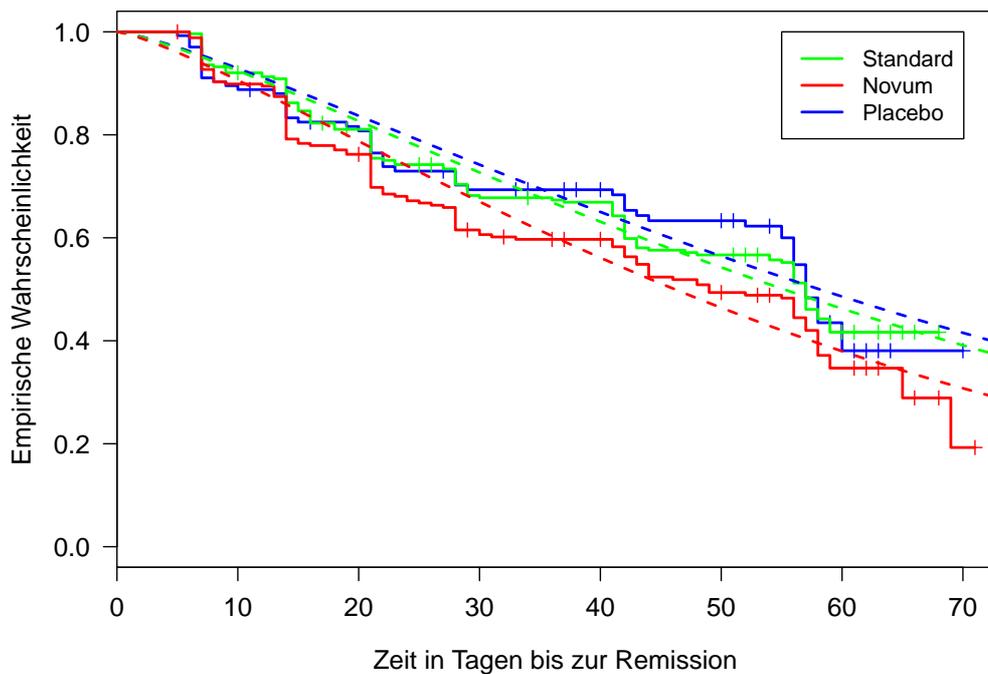


Abbildung 2.5: Aus den Daten geschätzte Kaplan-Meier Kurven (Treppenfunktionen) und mit Modell 2 angepasste Weibull-Verteilungen (—).

Die aus den Daten geschätzten Hazard Quotienten lauten: $HQ(\widehat{S}, P) = \frac{\hat{\lambda}_{S,2}}{\hat{\lambda}_{P,2}} = 1,07$ und $HQ(\widehat{N}, P) = \frac{\hat{\lambda}_{N,2}}{\hat{\lambda}_{P,2}} = 1,34$.

2.3 Zählprozesse in der Überlebenszeitanalyse

Die theoretische Grundlage für das Cox-Modell basiert auf dem Konzept der Zählprozesse. Dies ist naheliegend, da Modelle in der Überlebenszeitanalyse typischerweise für Daten entwickelt werden, die die Anzahl von verschiedenen Ereignissen zählen. Die Grundlagen hierfür stellen wir im nächsten Abschnitt vor. Im Folgenden führen wir die wesentlichen Begriffe ein, so dass wir unser Modell mit Hilfe von stochastischen Prozessen formulieren können. Für eine ausführliche Darstellung verweisen wir auf [Andersen et al. \[1993\]](#) oder auch [Fleming and Harrington \[2005\]](#).

Unerlässlich ist dabei zunächst der Begriff der Filtration, um die zu verschiedenen Zeitpunkten verfügbare Information darzustellen.

Definition 2.4 (Filtration) [vgl. Def. 1.2.4 [Fleming and Harrington, 2005](#), S.17]

Sei (Ω, \mathcal{F}, P) ein Wahrscheinlichkeitsraum und $\{\mathcal{F}_t : t \in \mathbb{B}\}$ eine Familie von σ -Unteralgebren der σ -Algebra \mathcal{F} . $\{\mathcal{F}_t : t \in \mathbb{B}\}$ heißt *Filtration*, falls $\mathcal{F}_s \subset \mathcal{F}_t$ für alle $s, t \in \mathbb{B}$ mit $s \leq t$ gilt. Sei $\mathcal{F}_{t+} := \sigma\{\bigcap_{h>0} \mathcal{F}_{t+h}\}$. Eine Filtration $\{\mathcal{F}_t : t \in \mathbb{B}\}$ heißt *rechtsstetig*, falls $\mathcal{F}_t = \mathcal{F}_{t+}$ für alle $t \in \mathbb{B}$ gilt.

Definition 2.5 (adaptiert) [vgl. Def. 1.2.5 [Fleming and Harrington, 2005](#), S.18]

Ein stochastischer Prozess $\{X(t) : t \in \mathbb{B}\}$ heißt *adaptiert* an die Filtration $\{\mathcal{F}_t : t \in \mathbb{B}\}$, falls $X(t)$ für alle $t \in \mathbb{B}$ ein \mathcal{F}_t -messbarer Prozess ist.

Gilt $\mathcal{F}_t = \sigma\{X(s); 0 \leq s \leq t\}$ für jedes $t \in \mathbb{B}$, so heißt $\{\mathcal{F}_t : t \in \mathbb{B}\}$ die von X erzeugte Filtration.

Offenbar ist ein stochastischer Prozess stets an seine erzeugte Filtration adaptiert. Die erzeugte Filtration ist die „kleinste“ Filtration, an die ein Prozess adaptiert ist (bei jeder anderen Filtration besitzen die σ -Algebren mehr Elemente).

Als nächstes führen wir den Begriff des Zählprozesses ein und zeigen, dass der folgende stochastische Prozess ein Zählprozess ist.

$$N_i(t) := \mathbb{1}_{\{\min\{T_i, \tilde{U}_i\} \leq t, \delta_i = 1\}} = \begin{cases} 1 : T_i \leq t \text{ (bei Person } i \text{ trat Ereignis bis Zeitpunkt } t \text{ ein)} \\ 0 : \text{sonst} \end{cases} \quad (2.5)$$

ist eine Treppenfunktion, die zu dem Zeitpunkt t , an dem bei Person i das Zielereignis eintritt, vom Funktionswert 0 auf 1 springt.

Definition 2.6 (Zählprozess) [vgl. Def. 1.2.6 [Fleming and Harrington, 2005](#), S.18]

Ein *Zählprozess* ist ein stochastischer Prozess $\{X(t) : t \in \mathbb{B}\}$, der adaptiert an die Filtration $\{\mathcal{F}_t : t \in \mathbb{B}\}$ ist, für den $X(0) = 0$ und $X(t) < \infty$ *f.s.* gilt und dessen Pfade mit Wahrscheinlichkeit 1 rechtsstetig sowie stückweise konstant mit Sprüngen der Größe 1 an den Unstetigkeitsstellen sind.

Nun führen wir einen weiteren Prozess ein, um schließlich alle in der Studie relevanten Informationen mittels stochastischer Prozesse ausdrücken zu können.

$$\begin{aligned} Y_i(t) &:= \mathbb{1}_{\{\min\{T_i, \tilde{U}_i\} \geq t\}} \\ &= \begin{cases} 1 & : \text{Person } i \text{ steht kurz vor dem Zeitpunkt } t \text{ unter Risiko} \\ 0 & : \text{sonst} \end{cases} \end{aligned} \quad (2.6)$$

In (N_i, Y_i, Z_i) sind für jede Person i alle relevanten Informationen erfasst (Z_i wie in (2.2) definiert). Die durch diese Prozesse erzeugte Filtration sei mit \mathcal{G}_t bezeichnet und hat folgende Form:

\mathcal{G}_t : Sei $\mathcal{G}_t^i := \sigma\{N_i(u), Y_i(u+), Z_i : 0 \leq u \leq t\}$ und sei $\mathcal{G}_t := \sigma\{\bigcup_{i=1}^n \mathcal{G}_t^i\}$ die kleinste σ -Algebra, die alle \mathcal{G}_t^i enthält für $i \in \{1, \dots, n\}$. Die so definierte Familie $\{\mathcal{G}_t : t \in \mathbb{B}\}$ ist eine rechtsstetige Filtration [vgl. Theorem 4.2.3 [Fleming and Harrington, 2005](#), S.135] und sei von nun an fixiert.

Der stochastische Prozess $\{N_i(t) : t \in \mathbb{B}\}$ ist ein Zählprozess, denn N_i ist adaptiert bezüglich \mathcal{G}_t , $N_i(0) = 0$, $N_i(t) < \infty$ und die Pfade sind rechtsstetig sowie stückweise konstant mit Sprüngen der Größe 1 an den Unstetigkeitsstellen.

Durch die Formulierung mittels Zählprozessen kann das Paar von Zufallsvariablen $(\min\{T_i, \tilde{U}_i\}, \delta_i)$ durch das Tripel (N_i, Y_i, Z_i) ersetzt werden.

Mit diesen Grundlagen sind wir nun in der Lage in Kapitel 3 eine Methode vorzustellen, mit der die Modell-Parameter β_k aus (2.1) geschätzt werden können und Aussagen über ihre asymptotischen Eigenschaften anzugeben.

3 Schätzer für die Modell-Parameter und ihre asymptotischen Eigenschaften

3.1 Schätzung der Modell-Parameter

Der Regressionsparameter β der Hazardfunktion aus (2.1) ist unter Berücksichtigung aller bisher getroffener Annahmen aus zensierten Daten zu schätzen, ohne dass die Form der Baseline-Hazardfunktion h_P bekannt ist. Als Schätzmethode verwenden wir ein Verfahren, das sich folgendermaßen aus der Maximum-Likelihood-Methode ableiten lässt:

Die Likelihoodfunktion kann in unserem Fall als Produkt zweier Faktoren so umgeformt werden, dass der erste Faktor alle wesentlichen Informationen, insbesondere den interessierenden Parameter β enthält und nicht von h_P abhängt. Der zweite Faktor hingegen hängt von h_P ab und liefert keine substantiellen Informationen zur Schätzung von β und wird weggelassen. Die so reduzierte Likelihoodfunktion wird *Partial-Likelihoodfunktion* genannt. Die Idee dieser Schätzmethode wurde von Cox entwickelt (vgl. Cox [1972] und Cox [1975]). Einzelheiten dazu sind im Anhang A.1 zu finden. In unserem Fall lautet die Partial-Likelihoodfunktion:

$$L(\beta) := \prod_{l=1}^L \frac{e^{\beta Z_{(l)}}}{\sum_{m \in R_{t_l}} e^{\beta Z_m}}, \quad (3.1)$$

$R_{t_l} := R_{T_{(l)}} := \{i \in \{1, \dots, n\} : \min\{T_i, \tilde{U}_i\} \geq T_{(l)}\}$ sei für $l \in \{1, \dots, L\}$ die Menge aller Personen, die bis kurz vor $T_{(l)}$ unter Risiko stehen.

Die gesuchte Lösung $\hat{\beta}$ heißt *Partial-Maximum-Likelihood-Schätzer*, maximiert die log-Partial-Likelihoodfunktion, kann mittels der Newton-Raphson Iteration ermittelt werden

und ist in Standardsoftware wie R und SAS implementiert.

Im nächsten Abschnitt werden wir Regularitätsbedingungen angeben, unter denen die Wahrscheinlichkeit, dass $\hat{\beta}$ die eindeutige Lösung des Gleichungssystems $\partial \log L(\beta) / \partial \beta = 0$ ist, gegen 1 strebt [vgl. Theorem VII.2.1. Andersen et al., 1993, S.497]. Mit Hilfe von Zählprozessen und Methoden aus der Martingal-Theorie ist es möglich, asymptotische Eigenschaften des Schätzers des Modellparameters β zu beweisen. Die Grundlage für diesen Ansatz stellt Aalen [1978] bereit. Speziell sehen wir, dass der Partial-Maximum-Likelihood-Schätzer konsistent und asymptotisch normalverteilt ist.

3.2 Asymptotische Eigenschaften des Schätzers des Modell-Parameters

Bevor wir Aussagen über die Konsistenz und die asymptotische Normalität des Partial-Maximum-Likelihood-Schätzers angeben, führen wir zunächst die dafür notwendigen Notationen und Regularitätsbedingungen ein.

Notation 3.1

Für einen Vektor $Y \in \mathbb{R}^d$ sei $\|Y\| := \max_{i \in \{1, \dots, d\}} |Y_i|$. Für eine Matrix $Y \in \mathbb{R}^{d \times m}$ sei $\|Y\| := \max_{i \in \{1, \dots, d\}; j \in \{1, \dots, m\}} |Y_{ij}|$. Für einen Vektor Y sei $|Y|$ die euklidische Norm. Für $t \in \mathbb{B}$ setze

$$\begin{aligned}
 S_n^{(0)}(\beta, t) &:= \frac{1}{n} \sum_{i=1}^n Y_i(t) e^{\beta Z_i}, \\
 S_n^{(1)}(\beta, t) &:= \frac{1}{n} \sum_{i=1}^n Z_i Y_i(t) e^{\beta Z_i} \stackrel{(2.2)}{=} \frac{1}{n} \sum_{i=1}^n \begin{pmatrix} Z_{iS} \\ Z_{iN} \end{pmatrix} Y_i(t) e^{\beta Z_i} \\
 &=: \begin{pmatrix} S_n^{(1,S)}(\beta, t) \\ S_n^{(1,N)}(\beta, t) \end{pmatrix}, \\
 S_n^{(2)}(\beta, t) &:= \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n Z_{iS} Y_i(t) e^{\beta Z_i} & 0 \\ 0 & \frac{1}{n} \sum_{i=1}^n Z_{iN} Y_i(t) e^{\beta Z_i} \end{pmatrix} \\
 &= \begin{pmatrix} S_n^{(1,S)}(\beta, t) & 0 \\ 0 & S_n^{(1,N)}(\beta, t) \end{pmatrix}.
 \end{aligned}$$

Regularitätsbedingungen 3.2 [vgl. Fleming and Harrington, 2005, S.289]

(R1) Es existiert $b > 0$, so dass $\int_0^b h_P(t)dt < \infty$.

(R2) Weiter sei \mathcal{B} eine kompakte Umgebung um β^0 und $s^{(j)}$ für $j = 0, 1, 2$ seien Funktionen, die auf $\mathcal{B} \times \mathcal{B}$ definiert sind und folgende Eigenschaft besitzen:

$$\sup_{t \in \mathcal{B}, \beta \in \mathcal{B}} \|S_n^{(j)}(\beta, t) - s^{(j)}(\beta, t)\| \xrightarrow[n \rightarrow \infty]{P} 0 \quad \text{für } j = 0, 1, 2,$$

d.h., dass die Folge $(S_n^{(j)})_{n \in \mathbb{N}}$ gleichmäßig gegen $s^{(j)}$ konvergiert für $j = 0, 1, 2$.

(R3) Die Funktionen $s^{(j)}$ für $j = 0, 1, 2$ seien beschränkt und es gebe eine Umgebung \mathcal{U} um 0, so dass für alle $t \in \mathcal{U}$ und $\beta \in \mathcal{B}$ gilt, $s^{(0)}(\beta, t) \neq 0$. Die Funktionen $s^{(j)}(\cdot, t)$ für $j = 0, 1, 2$ seien für $t \in \mathcal{B}$ gleichgradig stetig in β^0 .

(R4) Die wie folgt definierte Matrix sei positiv definit

$$\Sigma(\beta^0, b) := \int_0^b \left[\frac{s^{(2)}(\beta^0, t)}{s^{(0)}(\beta^0, t)} - \frac{s^{(1)}(\beta^0, t)}{s^{(0)}(\beta^0, t)} \left[\frac{s^{(1)}(\beta^0, t)}{s^{(0)}(\beta^0, t)} \right]^T \right] \cdot s^{(0)}(\beta^0, t) \cdot h_P(t) dt.$$

Eine Diskussion und Erläuterungen zur Bedeutung dieser Regularitätsbedingungen geben Andersen and Gill [1982].

Konsistenz

Satz 3.3 [vgl. Theorem 8.3.1 Fleming and Harrington, 2005, S.297]

Sei $\beta \in \mathcal{B}$ der Parameter im Cox-Modell und β^0 der wahre Wert von β . Sei $\hat{\beta}$ der Partial-Maximum-Likelihood-Schätzer von β . Dann ist $\hat{\beta}$ konsistent, d.h.

$$\hat{\beta} \xrightarrow[n \rightarrow \infty]{P} \beta^0.$$

□

Asymptotische Normalität

Satz 3.4 [vgl. Theorem 8.3.2 Fleming and Harrington, 2005, S.299]

Sei $\beta \in \mathcal{B}$ der Parameter im Cox-Modell und β^0 der wahre Wert von β . Sei $\hat{\beta}$ der Partial-Maximum-Likelihood-Schätzer von β . Sei $\Sigma(\beta^0, b)$ die in (R4) definierte Matrix und seien (R1)-(R4) erfüllt. Dann gilt:

$$\sqrt{n} (\hat{\beta} - \beta^0) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}_2 \left(0, [\Sigma(\beta^0, b)]^{-1} \right).$$

□

Die bisherigen Aussagen dieses Abschnittes wollen wir nun für unseren Fall verwenden. Wir geben die Form der Matrix $\Sigma(\beta^0, b)$ in unserem Modell an. Um zu zeigen, dass sie den Regularitätsbedingungen 3.2 genügt, benutzen wir den folgenden Satz:

Satz 3.5 [Andersen and Gill, 1982, nach Theorem 4.1]

Sei $\beta \in \mathcal{B}$ der Parameter im Cox-Modell und β^0 der wahre Wert von β . Sei $\hat{\beta}$ der Partial-Maximum-Likelihood-Schätzer von β . Seien die Tripel (N_i, Y_i, Z_i) für $i \in \{1, \dots, n\}$ unabhängig und identisch verteilt und $\{\mathcal{G}_t : t \in \mathbb{B}\}$ eine rechtsstetige Filtration, die die bis zur Zeit t zur Verfügung stehende Information darstellt. Dabei sei N_i wie in (2.5), Y_i wie in (2.6), Z_i wie in (2.2) und \mathcal{G}_t wie auf Seite 23 definiert. Sei h_i die Hazardfunktion im Cox-Modell aus (2.1) und sei $t \in \mathbb{B} = [0, b]$.

Dann sind die Regularitätsbedingungen 3.2 erfüllt, falls

- (i) $\int_0^b h_P(t) dt < \infty$,
- (ii) Y_i linksstetige Prozesse mit $P(Y_i(t) = 1) > 0$ sind und
- (iii) $\Sigma(\beta^0, b)$ positiv definit ist, wobei $s^{(j)}(\beta, t) = E \left[S_n^{(j)}(\beta, t) \right]$ für $j = 0, 1, 2$. □

Behauptung 3.6

Unter den in Kapitel 2 angegebenen Bezeichnungen und Annahmen erfüllt folgende Matrix die Bedingungen aus Satz 3.5

$$\Sigma(\beta^0, b) = \begin{pmatrix} \tilde{a} & \tilde{b} \\ \tilde{b} & \tilde{c} \end{pmatrix} \tag{3.2}$$

mit $\tilde{a} := w_S \int_0^b P(U_1 \geq t) f_S(t) dt - w_S^2 \int_0^b \frac{P(U_1 \geq t) f_S^2(t)}{f_S(t)w_S + f_N(t)w_N + f_P(t)w_P} dt$

$$\tilde{b} := -w_S w_N \int_0^b \frac{P(U_1 \geq t) f_S(t) f_N(t)}{f_S(t)w_S + f_N(t)w_N + f_P(t)w_P} dt$$

$$\tilde{c} := w_N \int_0^b P(U_1 \geq t) f_N(t) dt - w_N^2 \int_0^b \frac{P(U_1 \geq t) f_N^2(t)}{f_S(t)w_S + f_N(t)w_N + f_P(t)w_P} dt.$$

Dabei hängen die Dichten von β^0 ab, da

$$f_k(t) = h_k(t) S_k(t) = \begin{cases} h_P(t) e^{\beta_k^0} S_k(t) & \text{für } k = S, N \\ h_P(t) S_k(t) & \text{für } k = P. \end{cases}$$

Beweis:

Dass die Tripel (N_i, Y_i, Z_i) unabhängige und identisch verteilt sind, folgt direkt aus der Definition von T_i und U_i . Im Folgenden weisen wir die Voraussetzungen (i)-(iii) nach.

zu (i): $\int_0^b h_P(t)dt < \infty$, denn aus

$$\begin{aligned}
0 &< P(T_i \geq b) \cdot P(U_i \geq b) = P(T_i \geq b, U_i \geq b) \\
&= E[P(T_i \geq b \mid Z_i) \cdot P(U_i \geq b \mid Z_i)] \\
&= E\left[e^{-\exp(\beta Z_i) \int_0^b h_P(u)du} \cdot P(U_i \geq b \mid Z_i)\right] \\
&= E\left[\mathbb{1}_{\{e^{-\exp(\beta Z_i) \int_0^b h_P(u)du} > 0\}} e^{-\exp(\beta Z_i) \int_0^b h_P(u)du} \cdot P(U_i \geq b \mid Z_i)\right] \\
&\quad + \underbrace{E\left[\mathbb{1}_{\{e^{-\exp(\beta Z_i) \int_0^b h_P(u)du} = 0\}} e^{-\exp(\beta Z_i) \int_0^b h_P(u)du} \cdot P(U_i \geq b \mid Z_i)\right]}_{=0}
\end{aligned}$$

folgt $P\left(e^{-\exp(\beta Z_i) \int_0^b h_P(u)du} > 0\right) > 0$ und somit $\int_0^b h_P(u) du < \infty$.

zu (ii): Dass Y_i linksstetige Prozesse sind, folgt aus Kapitel 2.3. Den zweiten Teil der Aussage sieht man wie folgt

$$\begin{aligned}
P(Y_i(t) = 1) &\stackrel{(2.6)}{=} P(\mathbb{1}_{\{\min\{T_i, U_i, b\} \geq t\}} = 1) \\
&= P(T_i \geq t, U_i \geq t, b \geq t) \\
&= P(T_i \geq t) \cdot P(U_i \geq t) \cdot P(b \geq t) \\
&> 0.
\end{aligned}$$

zu (iii): Per Definition gilt (nach (R4) und Notation 3.1)

$$\Sigma(\beta^0, b) = \int_0^b \tilde{\Sigma}(\beta^0, t) \cdot h_P(t)dt$$

$$\text{mit } \tilde{\Sigma}(\beta^0, t) := \begin{pmatrix} s^{(1,S)}(\beta^0, t) - \frac{(s^{(1,S)}(\beta^0, t))^2}{s^{(0)}(\beta^0, t)} & -\frac{s^{(1,S)}(\beta^0, t)s^{(1,N)}(\beta^0, t)}{s^{(0)}(\beta^0, t)} \\ -\frac{s^{(1,S)}(\beta^0, t)s^{(1,N)}(\beta^0, t)}{s^{(0)}(\beta^0, t)} & s^{(1,N)}(\beta^0, t) - \frac{(s^{(1,N)}(\beta^0, t))^2}{s^{(0)}(\beta^0, t)} \end{pmatrix},$$

$$\text{und } s^{(0)}(\beta, t) = \frac{P(U_1 \geq t)}{h_P(t)} (f_S(t) w_S + f_N(t) w_N + f_P(t) w_P)$$

$$s^{(1,k)}(\beta, t) = \frac{P(U_1 \geq t)}{h_P(t)} f_k(t) w_k \quad \text{für } k = S, N$$

(vgl. Behauptung A.1).

Um zu zeigen, dass $\Sigma(\beta^0, b)$ positiv definit ist, zeigen wir, dass alle Hauptabschnittsdeterminanten größer 0 sind und nutzen aus, dass das Integral über positive Funktionen stets positiv ist.

1) Zeige, dass $\int_0^b \left[s^{(1,S)}(\beta^0, t) - \frac{(s^{(1,S)}(\beta^0, t))^2}{s^{(0)}(\beta^0, t)} \right] h_P(t) dt > 0$.

Da $0 < s^{(1,S)}(\beta^0, t) < s^{(0)}(\beta^0, t)$ für alle $t \in \mathbb{B}$ gilt, ist

$$\int_0^b \left[s^{(1,S)}(\beta^0, t) - \frac{(s^{(1,S)}(\beta^0, t))^2}{s^{(0)}(\beta^0, t)} \right] h_P(t) dt = \int_0^b s^{(1,S)}(\beta^0, t) \left[1 - \frac{s^{(1,S)}(\beta^0, t)}{s^{(0)}(\beta^0, t)} \right] h_P(t) dt > 0.$$

2) Zeige, dass

$$\int_0^b \left[s^{(1,S)}(\beta^0, t) - \frac{(s^{(1,S)}(\beta^0, t))^2}{s^{(0)}(\beta^0, t)} \right] h_P(t) dt \cdot \int_0^b \left[s^{(1,N)}(\beta^0, t) - \frac{(s^{(1,N)}(\beta^0, t))^2}{s^{(0)}(\beta^0, t)} \right] h_P(t) dt$$

$$- \left[\int_0^b \frac{s^{(1,S)}(\beta^0, t) s^{(1,N)}(\beta^0, t)}{s^{(0)}(\beta^0, t)} h_P(t) dt \right]^2 > 0.$$

Da $0 < s^{(1,S)}(\beta^0, t) + s^{(1,N)}(\beta^0, t) < s^{(0)}(\beta^0, t)$ für alle $t \in \mathbb{B}$ gilt, kann der linke Teil der obigen Ungleichung umgeformt und abgeschätzt werden zu

$$\int_0^b \frac{s^{(1,S)}(\beta^0, t) [s^{(0)}(\beta^0, t) - s^{(1,S)}(\beta^0, t)]}{s^{(0)}(\beta^0, t)} h_P(t) dt$$

$$\cdot \int_0^b \frac{s^{(1,N)}(\beta^0, t) [s^{(0)}(\beta^0, t) - s^{(1,N)}(\beta^0, t)]}{s^{(0)}(\beta^0, t)} h_P(t) dt$$

$$- \left[\int_0^b \frac{s^{(1,S)}(\beta^0, t) s^{(1,N)}(\beta^0, t)}{s^{(0)}(\beta^0, t)} h_P(t) dt \right]^2$$

$$> \left[\int_0^b \frac{s^{(1,S)}(\beta^0, t) s^{(1,N)}(\beta^0, t)}{s^{(0)}(\beta^0, t)} h_P(t) dt \right]^2 - \left[\int_0^b \frac{s^{(1,S)}(\beta^0, t) s^{(1,N)}(\beta^0, t)}{s^{(0)}(\beta^0, t)} h_P(t) dt \right]^2 = 0.$$

Somit ist $\Sigma(\beta^0, b)$ positiv definit. □

Da also $\Sigma(\beta^0, b)$ positiv definit ist, können wir die Matrix invertieren.

Die Matrix aus (3.2) kann im Allgemeinen nicht explizit berechnet werden, denn die Integrale können weder berechnet noch nach w_S , w_N und w_P aufgelöst werden, was allerdings für die Fallzahlplanung wünschenswert wäre. Deshalb geben wir im nächsten Abschnitt eine alternative Methode an, um die entsprechenden Einträge der Matrix näherungsweise zu erhalten. Hierbei orientieren wir uns an [Bernardo and Harrington \[2001\]](#).

3.3 Schätzer der Kovarianzmatrix

Definition 3.7 (Beobachtete Informationsmatrix) [nach [Casella and Berger, 1990](#), Kapitel 7.4]

Sei $\beta \in \mathcal{B}$ der Parameter im Cox-Modell und β^0 der wahre Wert von β . Sei $\hat{\beta}$ der Partial-Maximum-Likelihood-Schätzer von β . Sei L die Partial-Likelihoodfunktion aus (3.1). Dann heißt

$$\mathcal{J}(\hat{\beta}, b) := \frac{\partial^2}{\partial \beta^2} (-\log [L(\beta)]) \Big|_{\beta=\hat{\beta}} \quad (3.3)$$

die *beobachtete Informationsmatrix*.

Satz 3.8 [vgl. Theorem 8.2.1 (2) [Fleming and Harrington, 2005](#), S.290]

Unter den Bedingungen aus Satz 3.5 gilt für die Matrizen (3.2) und (3.3)

$$\sup_{t \in \mathbb{B}} \left\| \frac{1}{n} \mathcal{J}(\hat{\beta}, t) - \Sigma(\beta^0, t) \right\| \xrightarrow[n \rightarrow \infty]{P} 0.$$

□

Somit haben wir mit $1/n \mathcal{J}(\hat{\beta}, t)$ eine Matrix, mit der wir für hinreichend große n die Matrix $\Sigma(\beta^0, t)$ approximieren können.

Satz 3.9

Mit den in Kapitel 2 angegebenen Bezeichnungen ist

$$\mathcal{J}(\hat{\beta}, b) = \begin{pmatrix} \sum_{l=1}^L \frac{S_n^{(1,S)}(\hat{\beta}, t_l)}{S_n^{(0)}(\hat{\beta}, t_l)} - \left(\frac{S_n^{(1,S)}(\hat{\beta}, t_l)}{S_n^{(0)}(\hat{\beta}, t_l)} \right)^2 & \sum_{l=1}^L -\frac{S_n^{(1,S)}(\hat{\beta}, t_l) S_n^{(1,N)}(\hat{\beta}, t_l)}{[S_n^{(0)}(\hat{\beta}, t_l)]^2} \\ \sum_{l=1}^L -\frac{S_n^{(1,S)}(\hat{\beta}, t_l) S_n^{(1,N)}(\hat{\beta}, t_l)}{[S_n^{(0)}(\hat{\beta}, t_l)]^2} & \sum_{l=1}^L \frac{S_n^{(1,N)}(\hat{\beta}, t_l)}{S_n^{(0)}(\hat{\beta}, t_l)} - \left(\frac{S_n^{(1,N)}(\hat{\beta}, t_l)}{S_n^{(0)}(\hat{\beta}, t_l)} \right)^2 \end{pmatrix}.$$

Die Inverse existiert und lautet

$$\begin{aligned} [\mathcal{J}(\widehat{\beta}, b)]^{-1} &= \frac{1}{\sum_{l=1}^L \widehat{a}(t_l) \cdot \sum_{l=1}^L \widehat{c}(t_l) - \left(\sum_{l=1}^L \widehat{b}(t_l)\right)^2} \cdot \begin{pmatrix} \sum_{l=1}^L \widehat{c}(t_l) & -\sum_{l=1}^L \widehat{b}(t_l) \\ -\sum_{l=1}^L \widehat{b}(t_l) & \sum_{l=1}^L \widehat{a}(t_l) \end{pmatrix} \\ &=: \begin{pmatrix} \widehat{\text{Var}}(\widehat{\beta}_S) & \widehat{\text{Cov}}(\widehat{\beta}_S, \widehat{\beta}_N) \\ \widehat{\text{Cov}}(\widehat{\beta}_S, \widehat{\beta}_N) & \widehat{\text{Var}}(\widehat{\beta}_N) \end{pmatrix}, \end{aligned}$$

mit

$$\begin{aligned} \sum_{l=1}^L \widehat{a}(t_l) &= \sum_{l=1}^L \frac{S_n^{(1,S)}(\widehat{\beta}, t_l)}{S_n^{(0)}(\widehat{\beta}, t_l)} - \left(\frac{S_n^{(1,S)}(\widehat{\beta}, t_l)}{S_n^{(0)}(\widehat{\beta}, t_l)}\right)^2 \\ \sum_{l=1}^L \widehat{b}(t_l) &= \sum_{l=1}^L -\frac{S_n^{(1,S)}(\widehat{\beta}, t_l) S_n^{(1,N)}(\widehat{\beta}, t_l)}{\left[S_n^{(0)}(\widehat{\beta}, t_l)\right]^2} \\ \sum_{l=1}^L \widehat{c}(t_l) &= \sum_{l=1}^L \frac{S_n^{(1,N)}(\widehat{\beta}, t_l)}{S_n^{(0)}(\widehat{\beta}, t_l)} - \left(\frac{S_n^{(1,N)}(\widehat{\beta}, t_l)}{S_n^{(0)}(\widehat{\beta}, t_l)}\right)^2. \end{aligned}$$

Beweis:

Zunächst berechnen wir die Ableitungen der Partial-Likelihoodfunktion

$$\frac{\partial}{\partial \beta} \left(- \left[\frac{\partial}{\partial \beta} \log [L(\beta)] \right] \right) = \frac{\partial}{\partial \beta} [-U(\beta)],$$

dabei ist $U(\beta) := \frac{\partial L(\beta)}{\partial \beta}$ der score Vektor. Für ihn gilt folgendes:

$$\begin{aligned} U(\beta) &= \frac{\partial \log L(\beta)}{\partial \beta} \stackrel{(3.1)}{=} \frac{\partial}{\partial \beta} \left[\sum_{l=1}^L \beta Z_{(l)} - \log \left(\sum_{m \in R_{t_l}} e^{\beta Z_m} \right) \right] \\ &= \sum_{l=1}^L Z_{(l)} - \frac{\sum_{m \in R_{t_l}} Z_m e^{\beta Z_m}}{\sum_{m \in R_{t_l}} e^{\beta Z_m}} = \sum_{l=1}^L Z_{(l)} - \frac{\frac{1}{n} \sum_{i=1}^n Z_i e^{\beta Z_i} \mathbb{1}_{\{\min\{T_i, \tilde{U}_i\} \geq t_l\}}}{\frac{1}{n} \sum_{i=1}^n e^{\beta Z_i} \mathbb{1}_{\{\min\{T_i, \tilde{U}_i\} \geq t_l\}}} \\ &\stackrel{(2.6)}{=} \sum_{l=1}^L Z_{(l)} - \frac{S_n^{(1)}(\beta, t_l)}{S_n^{(0)}(\beta, t_l)}. \end{aligned}$$

Insgesamt erhalten wir nun für $J(\beta, b)$

$$J(\beta, b) = -\frac{\partial U(\beta)}{\partial \beta} = \sum_{l=1}^L \frac{\partial}{\partial \beta} \frac{S_n^{(1)}(\beta, t_l)}{S_n^{(0)}(\beta, t_l)} = \begin{pmatrix} \sum_{l=1}^L \frac{\partial \frac{S_n^{(1,S)}(\beta, t_l)}{S_n^{(0)}(\beta, t_l)}}{\partial \beta_S} & \sum_{l=1}^L \frac{\partial \frac{S_n^{(1,S)}(\beta, t_l)}{S_n^{(0)}(\beta, t_l)}}{\partial \beta_N} \\ \sum_{l=1}^L \frac{\partial \frac{S_n^{(1,N)}(\beta, t_l)}{S_n^{(0)}(\beta, t_l)}}{\partial \beta_S} & \sum_{l=1}^L \frac{\partial \frac{S_n^{(1,N)}(\beta, t_l)}{S_n^{(0)}(\beta, t_l)}}{\partial \beta_N} \end{pmatrix}$$

$$J(\hat{\beta}, b) \stackrel{\text{Beh. A.2}}{=} \begin{pmatrix} \sum_{l=1}^L \frac{S_n^{(1,S)}(\hat{\beta}, t_l)}{S_n^{(0)}(\hat{\beta}, t_l)} - \left(\frac{S_n^{(1,S)}(\hat{\beta}, t_l)}{S_n^{(0)}(\hat{\beta}, t_l)} \right)^2 & \sum_{l=1}^L -\frac{S_n^{(1,S)}(\hat{\beta}, t_l) S_n^{(1,N)}(\hat{\beta}, t_l)}{[S_n^{(0)}(\hat{\beta}, t_l)]^2} \\ \sum_{l=1}^L -\frac{S_n^{(1,S)}(\hat{\beta}, t_l) S_n^{(1,N)}(\hat{\beta}, t_l)}{[S_n^{(0)}(\hat{\beta}, t_l)]^2} & \sum_{l=1}^L \frac{S_n^{(1,N)}(\hat{\beta}, t_l)}{S_n^{(0)}(\hat{\beta}, t_l)} - \left(\frac{S_n^{(1,N)}(\hat{\beta}, t_l)}{S_n^{(0)}(\hat{\beta}, t_l)} \right)^2 \end{pmatrix}.$$

Um zu zeigen, dass die Matrix invertiert werden kann, zeigen wir, dass ihre Determinante ungleich 0 ist. Diese Rechnungen erfolgen analog zu 2) auf Seite 30 und sind hier deshalb nicht aufgeführt. \square

Die Matrix $J(\hat{\beta}, b)$ aus Satz 3.9 lässt sich weiter vereinfachen, indem wir $S_n^{(0)}(\beta, t)$, $S_n^{(1,S)}(\beta, t)$, $S_n^{(1,N)}(\beta, t)$ durch deren Erwartungswerte $E[S_n^{(0)}(\beta, t)]$, $E[S_n^{(1,S)}(\beta, t)]$, $E[S_n^{(1,N)}(\beta, t)]$ ersetzen. Simulationen zeigen, dass hierdurch in der Tat schon für die uns interessierenden Stichprobenumfänge die Abweichungen vernachlässigbar klein sind.

Somit erhalten wir zusammen mit Behauptung A.1:

$$J(\hat{\beta}, b) \approx \tilde{J}(\hat{\beta}, b) := \begin{pmatrix} \sum_{l=1}^L a(t_l) & \sum_{l=1}^L b(t_l) \\ \sum_{l=1}^L b(t_l) & \sum_{l=1}^L c(t_l) \end{pmatrix} \quad (3.4)$$

mit

$$a(t_l) := \frac{E[S_n^{(1,S)}(\hat{\beta}, t_l)]}{E[S_n^{(0)}(\hat{\beta}, t_l)]} - \left(\frac{E[S_n^{(1,S)}(\hat{\beta}, t_l)]}{E[S_n^{(0)}(\hat{\beta}, t_l)]} \right)^2$$

$$= \frac{f_S(t_l) \cdot w_S}{f_S(t_l) \cdot w_S + f_N(t_l) \cdot w_N + f_P(t_l) \cdot w_P} - \left(\frac{f_S(t_l) \cdot w_S}{f_S(t_l) \cdot w_S + f_N(t_l) \cdot w_N + f_P(t_l) \cdot w_P} \right)^2$$

$$b(t_l) := -\frac{E[S_n^{(1,S)}(\hat{\beta}, t_l)] E[S_n^{(1,N)}(\hat{\beta}, t_l)]}{\left(E[S_n^{(0)}(\hat{\beta}, t_l)] \right)^2} = -\frac{f_S(t_l) \cdot w_S \cdot f_N(t_l) \cdot w_N}{[f_S(t_l) \cdot w_S + f_N(t_l) \cdot w_N + f_P(t_l) \cdot w_P]^2}$$

$$\begin{aligned}
 c(t_l) &:= \frac{E \left[S_n^{(1,N)}(\widehat{\beta}, t_l) \right]}{E \left[S_n^{(0)}(\widehat{\beta}, t_l) \right]} - \left(\frac{E \left[S_n^{(1,N)}(\widehat{\beta}, t_l) \right]}{E \left[S_n^{(0)}(\widehat{\beta}, t_l) \right]} \right)^2 \\
 &= \frac{f_N(t_l) \cdot w_N}{f_S(t_l) \cdot w_S + f_N(t_l) \cdot w_N + f_P(t_l) \cdot w_P} - \left(\frac{f_N(t_l) \cdot w_N}{f_S(t_l) \cdot w_S + f_N(t_l) \cdot w_N + f_P(t_l) \cdot w_P} \right)^2.
 \end{aligned}$$

Satz 3.10

Die Inverse der Matrix (3.4) existiert und lautet:

$$\begin{aligned}
 \left[\widetilde{\mathcal{J}}(\widehat{\beta}, b) \right]^{-1} &= \frac{1}{\sum_{l=1}^L a(t_l) \cdot \sum_{l=1}^L c(t_l) - \left(\sum_{l=1}^L b(t_l) \right)^2} \cdot \begin{pmatrix} \sum_{l=1}^L c(t_l) & -\sum_{l=1}^L b(t_l) \\ -\sum_{l=1}^L b(t_l) & \sum_{l=1}^L a(t_l) \end{pmatrix} \\
 &=: \begin{pmatrix} \widetilde{\text{Var}}(\widehat{\beta}_S) & \widetilde{\text{Cov}}(\widehat{\beta}_S, \widehat{\beta}_N) \\ \widetilde{\text{Cov}}(\widehat{\beta}_S, \widehat{\beta}_N) & \widetilde{\text{Var}}(\widehat{\beta}_N) \end{pmatrix},
 \end{aligned}$$

wobei $a(t_l), b(t_l), c(t_l)$ wie in (3.4) definiert.

Beweis:

Um zu zeigen, dass die Matrix (3.4) invertiert werden kann, zeigen wir, dass ihre Determinante ungleich 0 ist.

$$\begin{aligned}
 \det \left(\widetilde{\mathcal{J}}(\widehat{\beta}, b) \right) &= \sum_{l=1}^L a(t_l) \cdot \sum_{l=1}^L c(t_l) - \left(\sum_{l=1}^L b(t_l) \right)^2 \\
 &= \sum_{l=1}^L \frac{f_S(t_l) \cdot w_S \cdot f_N(t_l) \cdot w_N}{(f_S(t_l) \cdot w_S + f_N(t_l) \cdot w_N + f_P(t_l) \cdot w_P)^2} \\
 &\quad \cdot \left[1 - \frac{f_S(t_l) w_S + f_N(t_l) w_N}{f_S(t_l) \cdot w_S + f_N(t_l) \cdot w_N + f_P(t_l) \cdot w_P} \right] \\
 &\quad + \sum_{l,m=1;l \neq m}^L \frac{f_S(t_l) \cdot w_S \cdot f_N(t_m) \cdot w_N}{d} \left[1 - \frac{e}{d} \right] \\
 &> 0
 \end{aligned}$$

mit

$$d := [f_S(t_l) \cdot w_S + f_N(t_l) \cdot w_N + f_P(t_l) \cdot w_P] [f_S(t_m) \cdot w_S + f_N(t_m) \cdot w_N + f_P(t_m) \cdot w_P]$$

$$e := f_S(t_l) \cdot w_S [f_S(t_m) \cdot w_S + f_N(t_m) \cdot w_N + f_P(t_m) \cdot w_P]$$

$$+ f_N(t_l) \cdot w_N [f_S(t_m) \cdot w_S + f_N(t_m) \cdot w_N + f_P(t_m) \cdot w_P].$$

Damit haben wir im Rahmen unseres Modells eine berechenbare Form für die Varianzen und Kovarianz der Partial-Maximum-Likelihood-Schätzer gefunden.

Depressionsstudie (Fortsetzung)

Die aus dem Datensatz mit der Partial-Maximum-Likelihood-Methode geschätzten Koeffizienten des Cox-Modells lauten $\widehat{\beta}_S = 0,058$ und $\widehat{\beta}_N = 0,264$. Daraus ergeben sich folgende Schätzwerte für die Hazard Quotienten $\widehat{HQ}(S, P) = 1,060$ und $\widehat{HQ}(N, P) = 1,303$ sowie folgende geschätzte Varianzen der Schätzer $\text{Var}(\widehat{\beta}_S) = 0,0264$ und $\text{Var}(\widehat{\beta}_N) = 0,0257$. Als Schätzwert für die Kovarianz der Schätzer ergibt sich $\text{Cov}(\widehat{\beta}_S, \widehat{\beta}_N) = 0,0182$.

Losgelöst von dem Weibull-Modell in dem Datenbeispiel auf Seite 18 ff. überprüfen wir hier direkt, ob bei der Depressionsstudie die Annahme proportionaler Hazardfunktionen gerechtfertigt ist. Hierzu gibt es in der Literatur zahlreiche Verfahren (vgl. etwa [Marubini and Valescchi \[1995\]](#), Kapitel 7). Wir stellen sowohl eine Methode basierend auf dem Likelihood-Quotienten-Test als auch eine graphische vor.

Im Cox-Modell haben unsere Hazardfunktionen nach (2.3) für jede Person $i \in \{1, \dots, n\}$ und $t \in \mathbb{B}$ die folgende Form:

$$h_i(t) = h_P(t) \cdot e^{\beta_S \cdot Z_{iS} + \beta_N \cdot Z_{iN}}.$$

Um die Annahme der proportionalen Hazardfunktionen im Cox-Modell zu überprüfen, nehmen wir zeitabhängige Variablen in das Modell mit auf. Die Hazardfunktionen haben dann folgende Form

$$h_i(t) = h_P(t) \cdot e^{\beta_S \cdot Z_{iS} + \beta_N \cdot Z_{iN} + \beta_{S,z} \cdot Z_{iS} \cdot t + \beta_{N,z} \cdot Z_{iN} \cdot t}. \quad (3.5)$$

Somit genügt das Modell aus (3.5) im Allgemeinen nicht der Annahme der proportionalen Hazardfunktionen. Nur in dem Fall, wenn $\beta_{S,z} = \beta_{N,z} = 0$, sind $HQ(S, P)$ und $HQ(N, P)$ im Zeitverlauf konstant. Ein Test, der die Hypothese $H_0: \beta_{S,z} = \beta_{N,z} = 0$ testet, ist ein Test auf proportionale Hazardfunktionen. Mit $\widehat{L}_{(2.3)}$ und mit $\widehat{L}_{(3.5)}$ sei der maximale Wert der Partial-Likelihoodfunktion von Modell (2.3) bzw. (3.5) bezeichnet. Die beiden Modelle

können nun im Rahmen des Likelihood-Quotienten-Tests anhand ihrer Werte der Partial-Likelihoodfunktionen miteinander verglichen werden. Als Niveau wählen wir 5%. Um die Hypothese H_0 zu testen, verwenden wir die Teststatistik

$$-2 \log \hat{L} = -2 \left(\log \hat{L}_{(2.3)} - \log \hat{L}_{(3.5)} \right) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \chi_2^2.$$

Der zum Wert $-2 \log \hat{L}_{(2.3)} + 2 \log \hat{L}_{(3.5)} = 0,23$ der Teststatistik gehörende p-Wert beträgt 89%. Daher kann die Hypothese zum 5% Niveau nicht verworfen werden. Insgesamt können wir also die beiden Modelle als gleichwertig ansehen und ziehen das Modell (2.3) dem Modell (3.5) vor. Graphisch kann die Annahme proportionaler Hazardfunktionen geprüft werden, indem die mit der Kaplan-Meier Methode geschätzten log-log-Überlebenszeitkurven $\log(-\log \hat{S}_k(t))$ der drei Medikamentengruppen $k = S, N, P$ gegen $\log(t)$ abgetragen werden. Ist die Annahme proportionaler Hazardfunktionen erfüllt, so verlaufen die Kurven parallel zueinander. In Abbildung 3.1 sind die entsprechenden Kurven gezeichnet.

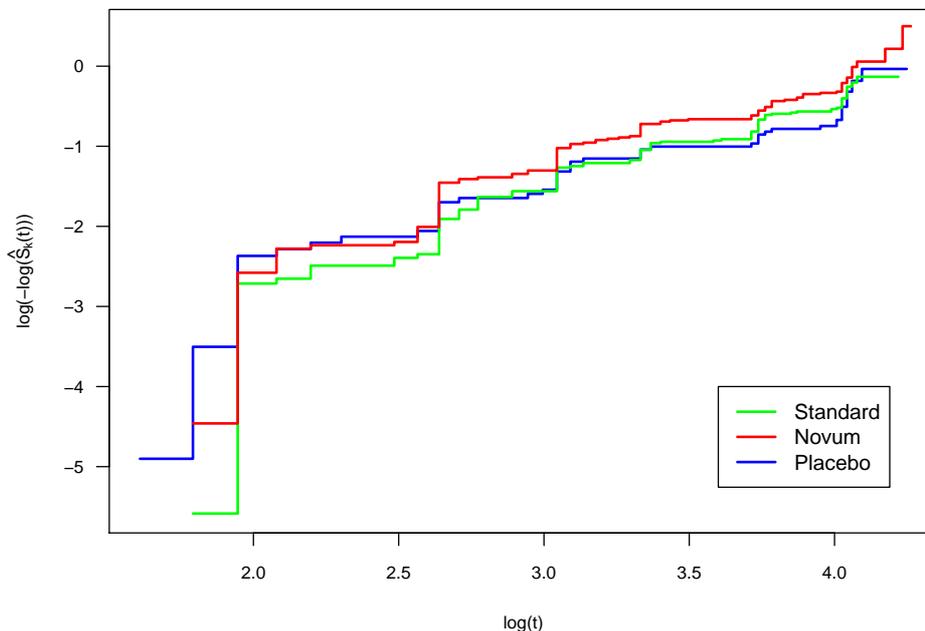


Abbildung 3.1: Graphischer Ansatz zur Prüfung der Annahme proportionaler Hazardfunktionen in der Depressionsstudie. Mit \hat{S}_k seien die nach der Kaplan-Meier Methode geschätzte Überlebenszeitkurven bezeichnet für $k = S, N, P$.

Insgesamt können wir somit in der Depressionsstudie annehmen, dass die Hazardfunktionen proportional zueinander sind. Zu demselben Ergebnis waren wir auch in dem Weibull-Modell auf Seite 18 ff. gekommen.

Unter Ausnutzung dieser Resultate bezüglich der Verteilung von $\hat{\beta}$ leiten wir im nächsten Kapitel ein semiparametrisches Testverfahren her, mit dem eine Nichtunterlegenheitsstudie im Gold Standard Design (vgl. Koch and Röhmel [2004]) ausgewertet werden kann.

4 Semiparametrisches Testverfahren

Sobald gegen eine Krankheit ein wirksames Medikament zugelassen ist, wird es in der Regel schwieriger ein noch wirksameres Präparat zu entwickeln. Ein Test, der rein eine bessere Wirkung von Novum untersucht, würde aufgrund der geringen Wirkungsunterschiede immens hohe Fallzahlen verlangen, was zu sogenannten mega trials führt. Wünschenswert ist aber nicht nur die Entwicklung rein auf die Krankheit bezogen besserer Medikamente, sondern auch die Entwicklung solcher, die ähnliche oder sogar klinisch nicht relevant unterlegene Wirkungserfolge aufweisen, dafür aber andere Vorzüge besitzen. Solche Vorteile können sein, dass sie weniger Nebenwirkungen aufweisen, besser verträglich sind, bei Patienten wirken, die gegen das Standardmedikament resistent sind, einfacher zu verabreichen sind oder weniger kosten.

Um assay sensitivity in einer Studie zu demonstrieren (d.h. zwischen wirksamen und unwirksamen Therapien unterscheiden zu können, vgl. etwa [Temple and Ellenberg \[2000\]](#), [Snapinn \[2004\]](#)), muss im Rahmen der Auswertung der klinischen Studie vorab getestet werden, ob Standard besser als Placebo wirkt. Alternativ kann zum Nachweis der assay sensitivity auch auf Überlegenheit des neuen Medikamentes über Placebo getestet werden. In diesem Kapitel stellen wir ein entsprechendes semiparametrisches Testverfahren vor.

Im Folgenden gehen wir ohne Einschränkung davon aus, dass eine Therapie um so besser wirkt, je höher der Wert der Hazardfunktion ist. Dies ist wünschenswert, wenn Heilung das Zielereignis ist. Der Fall, dass eine Therapie um so besser wirkt, je niedriger der Wert der Hazardfunktion ist (dies ist wünschenswert, wenn die Zeit bis zum Eintreten des Todes untersucht wird) kann analog behandelt werden. Um den therapeutischen Effekt zwischen zwei Medikamentengruppen zu quantifizieren, untersuchen wir die Therapieeffekte gemäß (2.4). Wir betrachten also den Hazard Quotienten auf der log-Skala.

4.1 Konzept der Überlegenheit (Vortest)

Vorab ist ein Überlegenheitstest durchzuführen, um zu gewährleisten, dass nur Medikamente im Markt eingeführt werden, die tatsächlich eine gewisse Wirkung aufweisen. Dieser Vortest untersucht die folgende Hypothese:

$$H_0^{Ueb,j} : \beta_j \leq 0 \quad \text{mit } j = S \text{ oder } j = N.$$

Welche Hypothese verwendet werden soll, muss a priori definiert werden. Für die Wahl von $H_0^{Ueb,j}$ wird in der Literatur sowohl $H_0^{Ueb,S} : \beta_S \leq 0$ (vgl. [Kieser and Friede \[2007\]](#), [Pigeot et al. \[2003\]](#), [Hasler et al. \[2008\]](#)) als auch $H_0^{Ueb,N} : \beta_N \leq 0$ (vgl. [Hauschke and Pigeot \[2005\]](#)) vorgeschlagen. Allerdings weist [Koch \[2005\]](#) darauf hin, dass gegen die Verwendung der ersten Hypothese folgendes spricht: Kann nicht gezeigt werden, dass das Standardmedikament dem Placebo überlegen ist, wird $H_0^{Ueb,S}$ nicht abgelehnt. In diesem Fall würde das neue Medikament wegen der Unwirksamkeit des Standardmedikamentes nicht auf dem Markt zugelassen, obwohl es wirksam sein kann.

Auch wir ziehen die Hypothese $H_0^{Ueb,N}$ der Hypothese $H_0^{Ueb,S}$ vor. Schließlich wird in diesem Fall Novum nicht durch die Unwirksamkeit des Standardmedikamentes „bestraft“, vielmehr hat es das neue Medikament „selbst in der Hand“ die geforderte Überlegenheit gegenüber Placebo zu zeigen, die Voraussetzung für die Anwendbarkeit des Nichtunterlegenheitstests ist. Außerdem ist es möglich, dass die Wirkung des Standardmedikamentes, die bei seiner Zulassung festgestellt wurde, heute gar nicht mehr genauso beobachtet worden wäre und somit eine nicht nachweisbare Überlegenheit in diesem Fall gar nicht auf fehlende assay sensitivity hinweisen muss. Da in der Literatur allerdings beide Hypothesen wahlweise verwendet werden, wollen wir hier der Vollständigkeit halber beide betrachten. Im Folgenden sei daher entweder $j = S$ oder $j = N$.

Als Teststatistik verwenden wir

$$T^{Ueb,j} := \frac{\hat{\beta}_j}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_j)}} = \sqrt{n_j + n_P} \frac{\hat{\beta}_j}{\hat{\sigma}_j}.$$

Die asymptotische Verteilung von $T^{Ueb,j}$ folgt aus der asymptotischen Normalität von $\hat{\beta}_j$ und dem Satz von Slutsky (vgl. beispielsweise [Ferguson \[1996\]](#), Kapitel 6). In Kapitel 4.2 werden wir die asymptotische Verteilung einer anderen Teststatistik bestimmen. Da die Verteilung von $T^{Ueb,j}$ analog dazu ermittelt werden kann, gehen wir an dieser Stelle nicht

naher darauf ein und stellen nur das Ergebnis vor:

$$T^{Ueb,j} \xrightarrow[n_j+n_P \rightarrow \infty]{\mathcal{D}} \mathcal{N} \left(\frac{\beta_j^0}{\sqrt{\text{Var}(\hat{\beta}_j)}}, 1 \right).$$

Also wird $H_0^{Ueb,j}$ verworfen, falls $T^{Ueb,j} > u_{1-\alpha}$ ist. Dabei bezeichnet $u_{1-\alpha}$ das $(1 - \alpha)$ -Quantil der Standardnormalverteilung. Ein Test dieser Bauart wird beispielsweise in [Bernardo and Harrington \[2001\]](#) betrachtet.

4.2 Konzept der Nichtunterlegenheit

Nichtunterlegenheit des neuen Medikamentes ist dann gezeigt, wenn ein bestimmter Anteil des Effekts des Standardmedikamentes gegenuber Placebo von dem neuen Medikament beibehalten wird. Die Unterlegenheit des neuen Medikamentes, die gerade noch als klinisch tolerabel angesehen werden kann, wird durch eine Nichtunterlegenheitsgrenze Δ quantifiziert. Sie muss vor Beginn der Studie festgelegt werden. Die Wahl von Δ hat einen groen Einfluss auf das Ergebnis der Auswertung. Bezuglich der Groenordnung dieser Grenze beschranken sich Empfehlungen und behordliche Richtlinien auf meist nur wenig konkrete Aussagen (vgl. hierzu beispielsweise den Entwurf der „Guidance for Industry on Non-Inferiority Clinical Trials“ der [FDA \[2010\]](#) oder die „Guideline on the choice of the non-inferiority margin“, [CHMP \[2005\]](#)).

Nichtunterlegenheitstest

Formulierung der Hypothese und ihrer Alternative

Wir erhalten folgendes Testproblem:

$$\begin{aligned} H_0^{Nu} : \log \left(\frac{h_N(t)}{h_P(t)} \right) &\leq \Delta \cdot \log \left(\frac{h_S(t)}{h_P(t)} \right) \\ H_1^{Nu} : \underbrace{\log \left(\frac{h_N(t)}{h_P(t)} \right)}_{=\beta_N} &> \Delta \cdot \underbrace{\log \left(\frac{h_S(t)}{h_P(t)} \right)}_{=\beta_S}. \end{aligned}$$

Der Term $\log (h_S(t)/h_P(t))$ wird *log-control-effect* und $\log (h_N(t)/h_P(t))$ wird *log-treatment-effect* genannt. Prinzipiell kann Δ aus dem Intervall $[0, \infty)$ gewahlt werden.

- Wählen wir $\Delta = 0$, so testen wir ausschließlich, ob N dem P überlegen ist,
- wählen wir $\Delta = 1$, so testen wir ausschließlich auf Überlegenheit von N bezüglich S ,
- wählen wir $\Delta > 1$, so testen wir auf sehr große Überlegenheit von N bezüglich S gemessen relativ zu Placebo.

Da wir eine 3-armige Nichtunterlegenheitsstudie betrachten, ergibt sich $\Delta \in (0, 1)$. Übliche Werte für Δ liegen nach [Lange and Freitag \[2005\]](#) oder auch [Hung et al. \[2005\]](#) im Bereich zwischen 0,5 und 0,8. Wir wollen untersuchen, ob Novum mindestens Δ -fach wirksamer ist als Standard (gemessen relativ zu Placebo und auf der log-Skala), d.h. mindestens $100 \cdot \Delta\%$ des log-control-effects beibehält (aus diesem Grund wird obige Hypothese im Englischen „retention of effect hypothesis“ genannt). Abbildung 4.1 stellt den Zusammenhang zwischen der Wahl von Δ und dem daraus resultierenden Bereich der Alternative dar. Die Linien entsprechen jeweils dem Rand der Hypothese, d.h. dort ist $\beta_N = \Delta \cdot \beta_S$. Je kleiner Δ gewählt wird, desto mehr Unterlegenheit des neuen Medikamentes wird erlaubt.

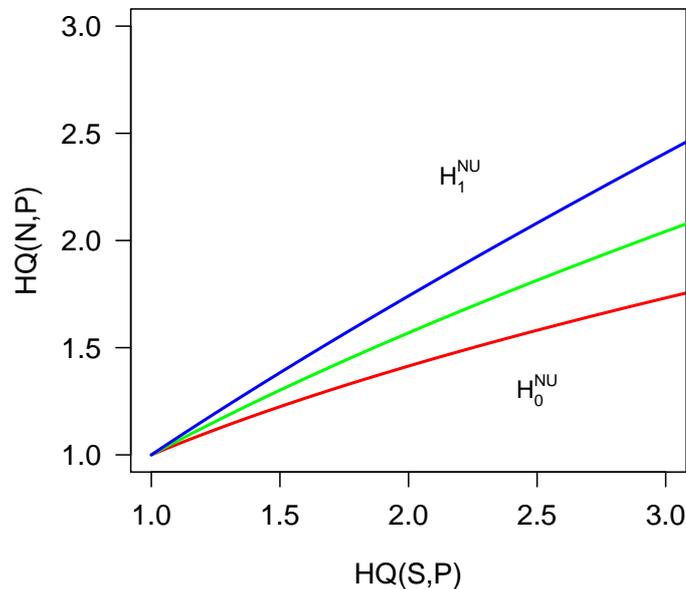


Abbildung 4.1: Bereich von Hypothese und Alternative für verschiedene Nichtunterlegenheitsgrenzen $\Delta \in \{0,5, 0,65, 0,8\}$, wobei $HQ(j, P)$ wie in (2.4) definiert ist für $j = S, N$.

Wahl der Teststatistik

Nun wollen wir eine Teststatistik entwickeln, um bei dem Hypothesentest die Entscheidung über Ablehnen oder Nichtablehnen der Hypothese treffen zu können. Um diese herzuleiten, schreiben wir die Nichtunterlegenheitshypothese um in

$$H_0^{Nu} : \beta_N - \Delta \cdot \beta_S \leq 0.$$

Wir schätzen $\beta_N - \Delta \cdot \beta_S$ durch $\widehat{\beta}_N - \Delta \cdot \widehat{\beta}_S$, wobei $\widehat{\beta}_N$ und $\widehat{\beta}_S$ die Partial-Maximum-Likelihood-Schätzer von β_N, β_S sind. Die Varianz von $\widehat{\beta}_N - \Delta \cdot \widehat{\beta}_S$,

$$\text{Var}(\widehat{\beta}_N - \Delta \cdot \widehat{\beta}_S) = \text{Var}(\widehat{\beta}_N) - 2 \cdot \Delta \cdot \text{Cov}(\widehat{\beta}_S, \widehat{\beta}_N) + \Delta^2 \cdot \text{Var}(\widehat{\beta}_S),$$

kann geschätzt werden durch

$$\widehat{\text{Var}}(\widehat{\beta}_N - \Delta \cdot \widehat{\beta}_S) = n(\widehat{\text{Var}}(\widehat{\beta}_N) - 2 \cdot \Delta \cdot \widehat{\text{Cov}}(\widehat{\beta}_S, \widehat{\beta}_N) + \Delta^2 \cdot \widehat{\text{Var}}(\widehat{\beta}_S)).$$

Somit erhalten wir folgende Teststatistik

$$\begin{aligned} T^{Nu} &= \sqrt{n} \frac{\widehat{\beta}_N - \Delta \cdot \widehat{\beta}_S}{\sqrt{\widehat{\text{Var}}(\widehat{\beta}_N - \Delta \cdot \widehat{\beta}_S)}} \\ &= \frac{\widehat{\beta}_N - \Delta \cdot \widehat{\beta}_S}{\sqrt{\widehat{\text{Var}}(\widehat{\beta}_N) - 2 \cdot \Delta \cdot \widehat{\text{Cov}}(\widehat{\beta}_S, \widehat{\beta}_N) + \Delta^2 \cdot \widehat{\text{Var}}(\widehat{\beta}_S)}}, \end{aligned}$$

deren asymptotische Verteilung mit Resultaten aus Kapitel 3 bestimmt werden kann, wie wir im nächsten Abschnitt zeigen werden.

Bestimmung des Verwerfungsbereichs

Als nächstes untersuchen wir, wann unsere Hypothese abgelehnt werden kann und bestimmen dazu den kritischen Bereich. Der kritische Bereich besteht nämlich aus den Werten der Teststatistik, die unter der Nullhypothese nur mit geringer Wahrscheinlichkeit auftreten. Hier sind unter der Hypothese kleine Werte für T^{Nu} zu erwarten. Die Bestimmung dieses Bereichs hängt sowohl von der Verteilung der Teststatistik als auch dem Signifikanzniveau α ab. Je nach Art der zu behandelnden Krankheit und Lebensbedrohlichkeit wird hierfür im Allgemeinen $\alpha = 2,5\%$ oder $\alpha = 5\%$ erlaubt.

Satz 4.1

Für die Verteilung der Teststatistik T^{Nu} gilt:

$$T^{Nu} = \frac{\widehat{\beta}_N - \Delta \cdot \widehat{\beta}_S}{\sqrt{\widehat{\text{Var}}(\widehat{\beta}_N) - 2 \cdot \Delta \cdot \widehat{\text{Cov}}(\widehat{\beta}_S, \widehat{\beta}_N) + \Delta^2 \cdot \widehat{\text{Var}}(\widehat{\beta}_S)}} \\ \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N} \left(\frac{\beta_N^0 - \Delta \cdot \beta_S^0}{\sqrt{\text{Var}(\widehat{\beta}_N) - 2 \cdot \Delta \cdot \text{Cov}(\widehat{\beta}_S, \widehat{\beta}_N) + \Delta^2 \cdot \text{Var}(\widehat{\beta}_S)}}, 1 \right).$$

Die asymptotische Verteilung der Teststatistik bestimmen wir mit Hilfe der im folgenden Satz vorgestellten δ -Methode:

Satz 4.2 (δ -Methode) [vgl. Theorem 3.1 [van der Vaart, 1998](#), S.26]

Sei T_n Schätzer von Θ und $\Psi : \mathbb{R}^d \rightarrow \mathbb{R}^m$ differenzierbar in Θ . Weiterhin gelte $\sqrt{n} (T_n - \Theta) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}_d(\mu, V^{-1})$. Dann gilt folgendes:

$$\sqrt{n} (\Psi(T_n) - \Psi(\Theta)) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}_m \left(\Psi'_{\Theta} \mu, \Psi'_{\Theta} V^{-1} (\Psi'_{\Theta})^T \right), \text{ wobei } \Psi'_{\Theta} := \left. \frac{\partial \Psi(x)}{\partial x} \right|_{x=\Theta}.$$

Beweis: (von Satz [4.1](#)) □

Nun wollen wir die δ -Methode in unserem Modell anwenden und setzen dafür

$$\begin{aligned} T_n &= \widehat{\beta} = (\widehat{\beta}_S, \widehat{\beta}_N), \\ \Theta &= \beta^0 = (\beta_S^0, \beta_N^0), \\ V &= \Sigma(\beta^0, b), \\ \mu &= (0, 0)^T, \\ \Psi \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= -\Delta \cdot x_1 + x_2 \text{ mit } \Psi'_{\beta^0} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (-\Delta, 1) \end{aligned}$$

und bemerken, dass nach Satz [3.4](#)

$$\sqrt{n}(\widehat{\beta} - \beta^0) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}_2(0, [\Sigma(\beta^0, b)]^{-1}) \text{ gilt.}$$

Zunächst stellen wir die einzelnen Ausdrücke des Satzes [4.2](#) in unserer Notation dar.

- $\sqrt{n} (\Psi(T_n) - \Psi(\Theta)) = \sqrt{n}(\Psi(\widehat{\beta}) - \Psi(\beta^0)) = \sqrt{n} [(\widehat{\beta}_N - \Delta \cdot \widehat{\beta}_S) - (\beta_N^0 - \Delta \cdot \beta_S^0)]$
- $\Psi'_{\Theta} \mu = (-\Delta, 1) \begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0$

$$\bullet \Psi'_{\Theta} V^{-1} (\Psi'_{\Theta})^T = \text{Var}(\widehat{\beta}_N) - 2 \cdot \Delta \cdot \text{Cov}(\widehat{\beta}_S, \widehat{\beta}_N) + \Delta^2 \cdot \text{Var}(\widehat{\beta}_S).$$

Zusammen gilt also nach Satz 4.2:

$$\begin{aligned} \sqrt{n} \left[(\widehat{\beta}_N - \Delta \cdot \widehat{\beta}_S) - (\beta_N^0 - \Delta \cdot \beta_S^0) \right] \\ \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, \text{Var}(\widehat{\beta}_N) - 2 \cdot \Delta \cdot \text{Cov}(\widehat{\beta}_S, \widehat{\beta}_N) + \Delta^2 \cdot \text{Var}(\widehat{\beta}_S)), \end{aligned}$$

was umgeformt zu folgendem Resultat führt

$$\sqrt{n} \frac{(\widehat{\beta}_N - \Delta \cdot \widehat{\beta}_S) - (\beta_N^0 - \Delta \cdot \beta_S^0)}{\sqrt{\text{Var}(\widehat{\beta}_N) - 2 \cdot \Delta \cdot \text{Cov}(\widehat{\beta}_S, \widehat{\beta}_N) + \Delta^2 \cdot \text{Var}(\widehat{\beta}_S)}} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, 1).$$

Mit dem Satz von Slutsky (vgl. beispielsweise [Ferguson \[1996\]](#), Kapitel 6) folgt die Behauptung. \square

Im nächsten Abschnitt simulieren wir den tatsächlichen Fehler 1. Art, um zu untersuchen, ab welchem Stichprobenumfang bereits eine Normalverteilung der Teststatistik angenommen werden kann.

Simulation des tatsächlichen Fehlers 1. Art

Ausgehend von Weibull-verteilten Überlebenszeiten untersuchen wir anhand von Simulationen das Verhalten des Tests für endliche Stichprobenumfänge und verschiedene Aufteilungen der Personen auf die Behandlungsgruppen. Als gesamten Stichprobenumfang beginnen wir mit $n = 30$ und verdoppeln die Gruppengrößen bis zu $n = 1.920$. Für die jeweiligen Patientenzahlen wurden verschiedene Verhältnisse für die Aufteilungen in die einzelnen Gruppen gewählt. Neben den Standardverhältnissen $n_S:n_N:n_P = 1:1:1$ und $2:2:1$ haben wir auch Verhältnisse untersucht, die von der Nichtunterlegenheitsgrenze $\Delta \in (0, 1)$ abhängen. Wählt man Δ nahe 1, so testet man schwerpunktmäßig N gegen S , d.h. in diesem Fall ist zu erwarten, dass ein Verhältnis, das bezogen auf N weniger Gewicht auf P und mehr auf S legt, das geeignete ist. Ist umgekehrt Δ klein, testet man stärker N gegen P , so dass bezogen auf N mehr Gewicht auf P und weniger auf S liegt. Daher wurde auch für verschiedene Δ Gruppenaufteilungen nach $\Delta:1:(1 - \Delta)$ untersucht: $1:2:1$ (für $\Delta = 0,5$), $13:20:7$ (für $\Delta = 0,65$) und $4:5:1$ (für $\Delta = 0,8$). Wir betrachten auch verschiedene Hazard Quotienten und zwar $HQ(S, P) \in \{1,25; 1,5; 3\}$. Den Skalen-Parameter λ_P setzen wir ohne Einschränkung auf $\lambda_P = 1$ und wählen λ_N gerade so, dass $\beta_N = \Delta \cdot \beta_S$

gilt. Der Form-Parameter $\vartheta \in \{0,5; 1; 1,5\}$ wird (wie bereits in Kapitel 2.2 beschrieben) für alle drei Überlebenszeitverteilungen gleich gewählt, um die Annahme proportionaler Hazardfunktionen zu gewährleisten. Als Niveau erlauben wir $\alpha = 2,5\%$ bzw. $\alpha = 5\%$. Für jede Parameterkonstellation wurden jeweils 10.000 Simulationen durchgeführt.

Um zu untersuchen, wie sich die Häufigkeitsverteilung der simulierten Werte der Teststatistik bei endlichem Stichprobenumfang verhält, haben wir in Abbildung 4.2 für einige Parameterwahlen die tailprobabilities dargestellt.

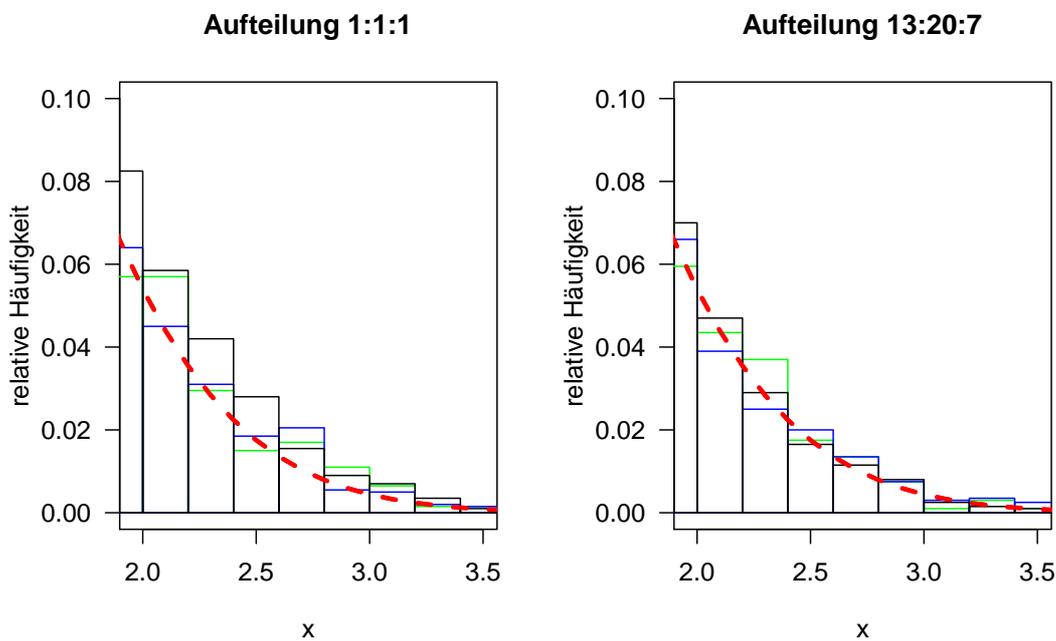


Abbildung 4.2: Tails der Histogramme der simulierten Werte der Teststatistik bei $n \in \{30, 120, 180\}$ und Dichte der Standardnormalverteilung (---) (Parameterwahl: $HQ = 1,5$, $\vartheta = 0,5$ und $\Delta = 0,65$).

Hier sieht man, wie sich die tailprobabilities für größer werdendes n bzw. Δ -abhängiger Aufteilung der Patienten in die Gruppen denen der Standardnormalverteilung gut annähern.

Als nächstes untersuchen wir, inwieweit sich die Wahl der Parameter der Überlebenszeitverteilungen auf die Einhaltung des Niveaus des Tests auswirkt. Die Nichtunterlegenheitsgrenze lassen wir fest, hier $\Delta = 0,65$. Der Übersicht halber haben wir die Ergebnisse für $\alpha = 5\%$ nur für $n \in \{120; 180; 480\}$ angegeben.

$HQ(S, P)$		$\vartheta = 0,5$			$\vartheta = 1$			$\vartheta = 1,5$		
		1,25	1,5	3	1,25	1,5	3	1,25	1,5	3
n	$n_S:n_N:n_P$									
30	1:1:1	3,41	3,67	3,32	3,48	3,32	3,41	3,80	3,54	3,20
	2:2:1	3,18	3,39	3,03	3,12	3,42	3,12	3,19	3,24	3,00
	13:20:7	2,57	2,59	2,60	2,69	2,84	2,61	2,69	2,68	2,64
60	1:1:1	3,21	3,11	3,26	3,18	3,26	2,86	3,08	3,03	3,23
	2:2:1	2,95	2,91	2,61	3,06	2,56	2,89	2,65	2,99	2,74
	13:20:7	2,50	2,88	2,47	2,97	2,48	2,55	2,81	2,28	2,50
120	1:1:1	3,00	2,89	2,66	2,52	2,75	2,74	2,60	2,80	2,85
		(5,53)	(5,48)	(5,59)	(5,88)	(5,52)	(5,22)	(5,36)	(5,45)	(5,23)
	2:2:1	2,52	2,73	2,73	2,69	2,71	2,62	2,65	2,77	2,62
		(5,30)	(5,17)	(5,14)	(5,39)	(5,20)	(5,33)	(5,22)	(5,08)	(5,05)
	13:20:7	2,45	2,47	2,28	2,44	2,33	2,48	2,59	2,70	2,76
		(4,92)	(4,72)	(4,81)	(5,15)	(4,93)	(4,91)	(4,59)	(4,60)	(4,77)
180	1:1:1	2,82	2,59	2,41	2,66	2,55	2,63	2,69	2,42	2,77
		(5,28)	(5,15)	(5,25)	(5,30)	(5,75)	(5,48)	(5,32)	(5,29)	(5,44)
	2:2:1	2,62	2,54	2,42	2,56	2,35	2,63	2,57	2,50	2,58
		(5,19)	(4,51)	(5,55)	(5,14)	(5,47)	(5,35)	(5,10)	(4,79)	(5,12)
	13:20:7	2,57	2,42	2,52	2,49	2,46	2,50	2,46	2,15	2,71
		(4,68)	(4,92)	(5,15)	(4,89)	(5,22)	(5,01)	(4,72)	(4,98)	(5,04)
240	1:1:1	2,76	2,90	2,65	2,53	2,59	2,46	2,43	2,53	2,61
		(5,15)	(5,41)	(4,99)	(5,02)	(5,49)	(4,94)	(5,32)	(5,14)	(5,47)
	2:2:1	2,45	2,55	2,53	2,51	2,62	2,52	2,57	2,50	2,43
		(5,38)	(5,27)	(5,08)	(5,12)	(5,25)	(5,13)	(5,59)	(5,49)	(5,27)
	13:20:7	2,34	2,52	2,42	2,67	2,49	2,49	2,45	2,65	2,57
		(4,78)	(4,99)	(4,42)	(4,59)	(4,94)	(5,10)	(5,08)	(4,84)	(4,80)
480	1:1:1	2,68	2,83	2,36	2,40	2,52	2,43	2,57	2,57	2,79
	2:2:1	2,46	2,60	2,35	2,71	2,62	2,36	2,59	2,45	2,20
	13:20:7	2,48	2,42	2,54	2,35	2,45	2,28	2,59	2,66	2,48
960	1:1:1	2,57	2,42	2,67	2,56	2,59	2,82	2,45	2,59	2,61
	2:2:1	2,67	2,60	2,46	2,48	2,52	2,55	2,73	2,31	2,62
	13:20:7	2,43	2,40	2,55	2,53	2,70	2,48	2,41	2,43	2,40
1.920	1:1:1	2,59	2,47	2,56	2,45	2,26	2,53	2,41	2,60	2,31
	2:2:1	2,60	2,36	2,38	2,46	2,51	2,46	2,29	2,45	2,71
	13:20:7	2,53	2,70	2,32	2,50	2,38	2,63	2,39	2,68	2,37

Tabelle 4.1: Simulierter tatsächlicher Fehler 1. Art in % zum nominalen Niveau von $\alpha = 2,5\%$ bzw. $\alpha = 5\%$ (Werte in Klammern) bei $\Delta = 0,65$.

Für kleine Stichprobenumfänge neigt der Test für alle Parameterkonstellationen dazu, sich anti-konservativ zu verhalten. Eine Abhängigkeit der Parameter der Überlebenszeitverteilungen auf die Niveaueinhaltung können wir nicht feststellen. Für $n \geq 180$ hält der Test für alle betrachteten Personenverhältnisse das Niveau zufriedenstellend ein.

Als nächstes untersuchen wir, inwieweit sich verschiedene Wahlen für Δ bei endlichen Stichprobenumfängen auf das Verhalten des Tests auswirken. Die Ergebnisse sind in Tabelle 4.2 abgebildet.

n	$n_S:n_N:n_P$	$\Delta=0,5$		$\Delta=0,65$		$\Delta=0,8$	
30	1:1:1	3,63		3,67		3,36	
	2:2:1	3,06		3,39		3,37	
	$\Delta:1:(1-\Delta)$	2,67		2,59		2,89	
60	1:1:1	3,22		3,11		3,23	
	2:2:1	2,92		2,91		2,58	
	$\Delta:1:(1-\Delta)$	2,64		2,49		2,79	
120	1:1:1	2,74	5,49	2,89	5,48	3,05	5,63
	2:2:1	2,85	4,86	2,73	5,17	2,96	5,18
	$\Delta:1:(1-\Delta)$	2,01	4,90	2,40	4,72	2,42	4,63
180	1:1:1	2,50	5,30	2,59	5,15	2,93	5,23
	2:2:1	2,72	5,20	2,54	4,51	2,56	5,32
	$\Delta:1:(1-\Delta)$	2,50	4,99	2,42	4,92	2,44	5,48
240	1:1:1	2,74	4,80	2,90	5,41	2,52	5,14
	2:2:1	2,66	5,23	2,48	5,27	2,39	5,28
	$\Delta:1:(1-\Delta)$	2,61	4,64	2,52	4,99	2,51	5,01
960	1:1:1	2,50		2,42		2,58	
	2:2:1	2,43		2,60		2,23	
	$\Delta:1:(1-\Delta)$	2,44		2,40		2,16	
1.920	1:1:1	2,39		2,47		2,62	
	2:2:1	2,63		2,36		2,70	
	$\Delta:1:(1-\Delta)$	2,44		2,70		2,36	

Tabelle 4.2: Simulierter tatsächlicher Fehler 1. Art in % zum nominalen Niveau von $\alpha = 2,5\%$ (links) und $\alpha = 5\%$ (rechts) für $HQ(S, P) = 1, 5$ und $\vartheta = 0,5$.

Auch hier hält der Test ab $n = 180$ für alle Parameterkonstellationen das Niveau zufriedenstellend ein. Eine Abhängigkeit dieser Ergebnisse von Δ scheint nicht gegeben zu sein. Zusammenfassend stellen wir fest, dass für alle Parameterkonstellationen ab $n = 180$ der Test das nominale Niveau zufriedenstellend einhält. Für an Δ angepasste Aufteilungen der Personen sind hierfür bereits $n = 120$ Beobachtungen ausreichend.

Beinhaltet die Studie zusätzlich auch zensierte Beobachtungen, so sind die Fallzahlen entsprechend zu erhöhen, damit auch dann der Test das nominale Niveau zufriedenstellend einhält.

4.3 Auswertung der Depressionsstudie

Da die Voraussetzung der proportionalen Hazardfunktionen angenommen werden kann (vgl. Ergebnisse auf Seite 35 ff.), können wir das Testverfahren schließlich anwenden. Insgesamt ist also folgende Hypothese zu betrachten:

$$H_0 : H_0^{Ueb,j} \cup H_0^{Nu} = \{\beta_j \leq 0\} \cup \{\beta_N \leq \Delta \cdot \beta_S\} \quad \text{für } j = S \text{ oder } j = N.$$

Wie bereits beschrieben, wird H_0 abgelehnt, falls sowohl $H_0^{Ueb,j}$ als auch H_0^{Nu} abgelehnt werden können. Aufgrund der hierarchischen Struktur dieses Testproblems wird H_0 zum Niveau α abgelehnt, falls $H_0^{Ueb,j}$ und H_0^{Nu} jeweils zum Niveau α abgelehnt werden können. Der zugehörige Test heißt *Intersection-Union-Test*.

$H_0^{Ueb,S}$: Zunächst schätzen wir den Parameter des Cox-Modells und seine Varianz aus den Daten und erhalten folgende Ergebnisse:

$$\hat{\beta}_S = 0,054 \quad \text{mit} \quad \widehat{\text{Var}}(\hat{\beta}_S) = 0,026.$$

Als nächstes testen wir die Überlegenheitshypothese und erhalten den p-Wert 37,09%.

$H_0^{Ueb,N}$: Analoges Vorgehen wie oben ergibt

$$\hat{\beta}_N = 0,261 \quad \text{mit} \quad \widehat{\text{Var}}(\hat{\beta}_N) = 0,026$$

und einen p-Wert von 5,17%.

Weder zum Niveau von 5% noch zum 2,5% Niveau könnte also basierend auf diesem Testergebnis gezeigt werden, dass das Standardmedikament besser als Placebo wirkt. Auch für das neue Medikament könnte dies anhand der uns zugrunde liegenden Daten weder zum Niveau von 5% noch zum 2,5% Niveau gezeigt werden. Deshalb müssen wir das Testverfahren an dieser Stelle abbrechen unabhängig davon, welcher Vortest ausgewählt worden wäre.

5 Konzept der Fallzahlplanung

Wie bereits in Kapitel 2 erläutert, ist vor der Durchführung einer klinischen Studie die Anzahl der teilnehmenden Personen, ihre Aufteilung in die Gruppen und die Dauer der Studie, d.h. die Anzahl der Beobachtungsperioden festzulegen. Bezüglich der Planung des Stichprobenumfangs ist darauf zu achten, dass aus ethischer Sicht und aus Kostengründen ein unnötig großer Stichprobenumfang zu vermeiden ist. Jedoch dürfen die Fallzahlen nicht zu klein gewählt werden, da die Daten dann möglicherweise zu wenig Informationen beinhalten, um einen klinisch relevanten Unterschied zwischen den Behandlungen nachweisen zu können.

Die erforderliche Anzahl an Ereignissen bestimmen wir deshalb gerade so, dass der Nichtunterlegenheitstest zum Niveau α eine tatsächliche Abweichung von der Hypothese H_0^{Nu} mit einer bestimmten Wahrscheinlichkeit γ (der sogenannten Güte) aufdeckt. Typische Werte sind $\alpha = 2,5\%$ (bzw. $\alpha = 5\%$) und $\gamma = 80\%$. Danach wird daraus die Anzahl an Personen insgesamt bestimmt. Dieses zweistufige Verfahren ist erforderlich, da die Güte von der Anzahl der Ereignisse abhängt anstatt von der gesamten Personenanzahl.

Da die Güte des Intersection-Union Tests kleiner oder gleich der Güte des Nichtunterlegenheitstests ist, müssten wir unsere Berechnungen eigentlich basierend auf der Güte des Intersection-Union Tests durchführen. Diese Güte ist im Allgemeinen jedoch nicht analytisch zu berechnen. Unser Vorgehen liefert auch für den Intersection-Union Test passende Fallzahlen, denn in Kapitel 6 können wir folgendes zeigen: planen wir die Fallzahlen unter der üblicherweise verwendeten Alternative $\beta_S = \beta_N \neq 0$ so, dass wir beim Nichtunterlegenheitstest eine Güte von 80% erreichen, führt dies beim Intersection-Union Tests zu einer nur unwesentlich geringeren Güte, da die Güte des Überlegenheitstests in diesem Fall nahe 1 ist.

5.1 Berechnung der insgesamt erforderlichen Ereignisse

Die für die Fallzahlplanung relevante asymptotische Güte ist nach Satz 4.1 und Abschnitt 3.3 gegeben durch

$$\begin{aligned} \gamma &= P(T^{Nu} > u_{1-\alpha}) \\ &= 1 - \Phi \left(u_{1-\alpha} - \frac{\beta_N - \Delta \cdot \beta_S}{\sqrt{\widetilde{\text{Var}}(\widehat{\beta}_N) - 2 \cdot \Delta \cdot \widetilde{\text{Cov}}(\widehat{\beta}_S, \widehat{\beta}_N) + \Delta^2 \cdot \widetilde{\text{Var}}(\widehat{\beta}_S)}} \right) + o_p(1), \end{aligned} \quad (5.1)$$

wobei Φ die Verteilungsfunktion der Standardnormalverteilung und o_p das stochastische Landau-Symbol (vgl. beispielsweise van der Vaart [1998], Kapitel 2.2) ist. Umformungen liefern die Gleichung, mit der wir im Folgenden arbeiten werden:

$$\left(\frac{\beta_N - \Delta \cdot \beta_S}{u_{1-\gamma} - u_{1-\alpha}} \right)^2 \approx \widetilde{\text{Var}}(\widehat{\beta}_N) - 2 \cdot \Delta \cdot \widetilde{\text{Cov}}(\widehat{\beta}_S, \widehat{\beta}_N) + \Delta^2 \cdot \widetilde{\text{Var}}(\widehat{\beta}_S). \quad (5.2)$$

Für die Planung der Fallzahlen werden Parameter gewählt, die zur Alternative gehören. Da wir von ungefähr gleicher Wirkung der zu testenden Medikamente ausgehen, können wir vereinfachen und $\beta_S = \beta_N \neq 0$ setzen. Wäre ein Medikament deutlich besser als das andere, wäre eine Überlegenheitsstudie das geeignete Verfahren. Wird allerdings ein vernachlässigbarer Wirkungseffekt zwischen den Medikamenten erwartet, bietet sich eine Überlegenheitsstudie nicht an, da hier sehr hohe Fallzahlen benötigt werden, was zu sogenannten mega trials führt. Unter dieser Annahme gilt dann für die Hazardfunktionen $h_S = h_N =: h$, für die Überlebenszeitfunktionen $S_S = S_N =: S$ und für die Dichten $f_S = f_N =: f$. Mit dieser Vereinfachung erhalten wir

$$\widetilde{\text{Var}}(\widehat{\beta}_N) - 2 \cdot \Delta \cdot \widetilde{\text{Cov}}(\widehat{\beta}_S, \widehat{\beta}_N) + \Delta^2 \cdot \widetilde{\text{Var}}(\widehat{\beta}_S) = g(\theta, \theta_P, w, L, t) \quad (5.3)$$

mit

$$g(\theta, \theta_P, w, L, t) \quad (5.4)$$

$$:= \frac{\sum_{l=1}^L \frac{f^\theta(t_l) w_S [f^\theta(t_l) w_N + f_P^{\theta_P}(t_l) w_P]}{\text{Nenner}} - \frac{2\Delta [f^\theta(t_l)]^2 w_S w_N}{\text{Nenner}} + \Delta^2 \frac{f^\theta(t_l) w_N [f^\theta(t_l) w_S + f_P^{\theta_P}(t_l) w_P]}{\text{Nenner}}}{\left[\sum_{l=1}^L \frac{f^\theta(t_l) w_S [f^\theta(t_l) w_N + f_P^{\theta_P}(t_l) w_P]}{\text{Nenner}} \right] \left[\sum_{l=1}^L \frac{f^\theta(t_l) w_N [f^\theta(t_l) w_S + f_P^{\theta_P}(t_l) w_P]}{\text{Nenner}} \right] - \left[\sum_{l=1}^L \frac{[f^\theta(t_l)]^2 w_S w_N}{\text{Nenner}} \right]^2}$$

wobei

θ der Parameter der Dichte f ,

θ_P der Parameter der Dichte f_P ,

$w := (w_S, w_N, w_P)$,

$t := (t_1, \dots, t_L)$ und

Nenner $:= [f(t_l)(w_S + w_N) + f_P(t_l)w_P]^2$ sei.

Die Approximation für die asymptotischen Varianzen und die asymptotische Kovarianz der Partial-Maximum-Likelihood-Schätzer hängt also von den in der Studie beobachteten Ereigniszeitpunkten t_1, \dots, t_L ab. Um die Größe des erforderlichen Stichprobenumfanges planen zu können, sind also vorab Annahmen über die Lage der Ereigniszeitpunkte zu treffen. Anschließend kann dann die Anzahl der insgesamt benötigten Ereignisse L in Abhängigkeit von w_S, w_N, w_P und den Dichten f, f_P berechnet werden.

5.2 Berechnung der insgesamt erforderlichen Personen

Erwarten wir keine zensierten Beobachtungen, so entspricht die Anzahl der Ereignisse L dem insgesamt benötigten Stichprobenumfang n . Nun stellen wir den Zusammenhang zwischen L und n vor, wenn zusätzlich zensierte Beobachtungen berücksichtigt werden sollen. Es gilt

$$L = n \cdot \psi, \tag{5.5}$$

wobei

$$\psi := P(T_S < b, T_S < U_1) w_S + P(T_N < b, T_N < U_1) w_N + P(T_P < b, T_P < U_1) w_P$$

die Wahrscheinlichkeit für ein Ereignis im Verlauf der Studie ist.

Mit diesen Ergebnissen können wir nun einen Algorithmus zur Bestimmung der Fallzahl aufstellen. Wegen der Bedeutung und des Umfangs haben wir hierfür ein neues Kapitel vorgesehen.

6 Algorithmen zur Bestimmung der Fallzahl

In diesem Kapitel werden Algorithmen entwickelt, die es ermöglichen, für unser Modell unter Einbeziehung der Ergebnisse der vorangegangenen Kapitel, geeignete Fallzahlen pro Patientengruppe numerisch zu bestimmen. Wie wir bereits in Kapitel 5 gesehen haben, geschieht dies indem - bei vorgegebenen Parameterwahlen für das Niveau, die Nichtunterlegenheitsgrenze, die Wirkung der Medikamente zueinander und die Anteile an zensierten Beobachtungen - der Stichprobenumfang gerade so gewählt wird, dass eine gewünschte Güte erreicht wird. Dabei wollen wir die Anzahl der Personen pro Behandlungsgruppe gerade so wählen, dass so wenig Personen wie möglich an der Studie teilnehmen müssen.

Um für die Algorithmen einen guten Startwert wählen zu können und außerdem einen Einblick in die Verhältnisse der Gruppengrößen zueinander zu erhalten, interessiert uns zunächst nicht, zu welchen Zeitpunkten während der Studie genau die Ereignisse eintreten, sondern nur, ob die Ereignisse eintreten oder nicht (binäre Endpunkte). Vereinfachend nehmen wir an, dass alle Ereignisse an demselben Zeitpunkt eintreten. Nach diesen Vorüberlegungen kehren wir schließlich zu unserem Modell zurück und bestimmen konkrete Fallzahlen. Schließlich untersuchen wir die Güte für die in Kapitel 4 vorgestellten Testverfahren mittels Simulationen bezüglich der mit den Algorithmen berechneten Stichprobenumfängen und diskutieren die Robustheit des Nichtunterlegenheitstests bei Abweichungen von der Annahme proportionaler Hazardfunktionen.

6.1 Berechnung der Lage der Ereigniszeitpunkte und ihre insgesamt erforderliche Anzahl

Die Lage der Ereigniszeitpunkte t_l ergibt sich aus der jeweiligen Weibull-Verteilung. Unter dieser Annahme bestimmen wir nun für Gruppe k die Lage der L_k nach der Größe geordneten Ereigniszeitpunkte t_l^k folgendermaßen: t_l^k ist die Lösung der Gleichung

$$L_k \cdot P(T_k > x) = L_k - l$$

in x für $l = 1, \dots, L_k - 1$ und $k = S, N, P$. Umgeformt nach t_l^k können wir die Ereigniszeitpunkte mittels folgender Formel berechnen

$$t_l^k = \sqrt[\vartheta]{-\frac{1}{\lambda_k} \cdot \log\left(1 - \frac{l}{L_k}\right)}. \quad (6.1)$$

Den letzten Ereigniszeitpunkt in jeder Gruppe können wir mit dieser Methode nicht bestimmen, da $L_k \cdot P(T_k > x) = 0$ keine reelle Lösung hat (der Bildbereich von $T_k \sim Wb(\lambda_k, \vartheta)$ ist nämlich $\mathbb{R}_{>0}$). Deshalb setzen wir den letzten Ereigniszeitpunkt auf das Studienende, also $t_{L_k}^k = b$ für $k = S, N, P$.

Es ist darauf zu achten, dass t_l^k für alle $l \in \{1, \dots, L_k - 1\}$ und $k = S, N, P$ im Beobachtungszeitraum $\mathbb{B} = [0, b]$ liegt. Falls $\max\{t_l^k : l \in \{1, \dots, L_k - 1\}\} > b$, so fügen wir $j \in \mathbb{N}$ zusätzliche Ereigniszeitpunkte hinzu und bestimmen die Lage der $L_k + j - 1$ Ereigniszeiten nach der oben vorgestellten Methode. Dabei wird j genau so gewählt, dass bei insgesamt $L_k + j - 2$ Ereignissen für die Ereigniszeit $t_{L_k-1}^k > b$ aber bei insgesamt $L_k + j - 1$ Ereignissen $t_{L_k-1}^k \leq b$ gilt. Für unsere Berechnungen betrachten wir dann diese ersten $L_k - 1$ Ereigniszeitpunkte.

Falls keine zensierten Beobachtungen berücksichtigt werden sollen ist $n_k = L_k$. Wenn zusätzlich auch zensierte Beobachtungen erwartet werden, stehen diese beiden Größen nach Formel (5.5) in folgendem Zusammenhang

$$L_k = n \cdot w_k \cdot P(T_k < b, T_k < U_1). \quad (6.2)$$

Neben der Verteilung der Überlebenszeiten nehmen wir auch eine Verteilung für die Zensierungszeiten an.

Wahl der Zensierungsverteilung und ihrer Parameter

Bezüglich der Wahl der Zensierungsverteilung ist folgendes zu beachten: Nehmen wir an, dass die Zensierungszeiten im Verlauf der Studie alle mit derselben Rate eintreten, ist eine Exponentialverteilung (Weibull-Verteilung mit Form-Parameter 1) als Zensierungsverteilung zu wählen (vgl. Anhang A.3). Wird hingegen vermutet, dass die Zensierungszeiten mit einer wachsenden oder fallenden Rate eintreten, wäre eine Weibull-Verteilung mit Form-Parameter ungleich 1 die geeignete Wahl. Hierbei ist der Form-Parameter größer als 1 zu wählen, wenn eine wachsende Rate angenommen wird, für eine fallende ist er kleiner als 1 zu wählen. Wie bereits erwähnt, nehmen wir an, dass die Verteilung der Zensierungszeiten aller Personen gleich ist. Außerdem wählen wir den Form-Parameter der Überlebenszeitverteilungen und der Zensierungsverteilung gleich. Somit ist

$$\begin{aligned} T_S, T_N &\sim Wb(\lambda, \vartheta), \\ T_P &\sim Wb(\lambda_P, \vartheta) \text{ und} \\ U_i &\sim Wb(u, \vartheta) \quad \text{für } i \in \{1, \dots, n\}. \end{aligned}$$

Der Skalen-Parameter u der Zensierungsverteilung wird so gewählt, dass die Wahrscheinlichkeit für eine zufällig zensierte Beobachtung einer Person aus der Standardgruppe einen bestimmten Wert annimmt, dieser sei mit p_z^S bezeichnet.

$$\begin{aligned} P(U_1 < T_S, U_1 < b) &= \int_{\mathbb{R}} f_{U_1}(x) P(x < \min\{T_S, b\}) dx = \int_0^b f_{U_1}(x) P(T_S > x) dx \\ &= \int_0^b u \vartheta x^{\vartheta-1} e^{-u x^\vartheta} e^{-\lambda x^\vartheta} dx = -u \frac{e^{-b^\vartheta(\lambda+u)} - 1}{\lambda + u} =: p_z^S \end{aligned} \quad (6.3)$$

Diese Wahrscheinlichkeit hängt also von der Studiendauer b ab. Die Dauer wiederum hängt von der Wahrscheinlichkeit einer fest zensierten Beobachtung in der Standardgruppe ab, die wir mit p_f^S bezeichnen:

$$P(b < T_S, b < U_1) = P(b < T_S) P(b < U_1) = e^{-b^\vartheta(\lambda+u)} =: p_f^S.$$

Nach der Studiendauer aufgelöst, erhalten wir daraus

$$b = \left(-\frac{\log(p_f^S)}{\lambda + u} \right)^{1/\vartheta} \quad \text{für } p_f^S > 0. \quad (6.4)$$

In Abbildung 6.1 ist die Abhängigkeit zwischen Studiendauer b und Anteil der fest zensierten Patienten aus der Standardgruppe p_f^S nach (6.4) für verschiedene Wahlen der Skalen- und Form-Parameter und Anteile von zufällig zensierten Beobachtungen in der Standardgruppe p_z^S dargestellt. Je nachdem ob die Beobachtungen täglich, wöchentlich oder monatlich durchgeführt werden, betragen die Zeiteinheiten Tage, Wochen oder Monate.

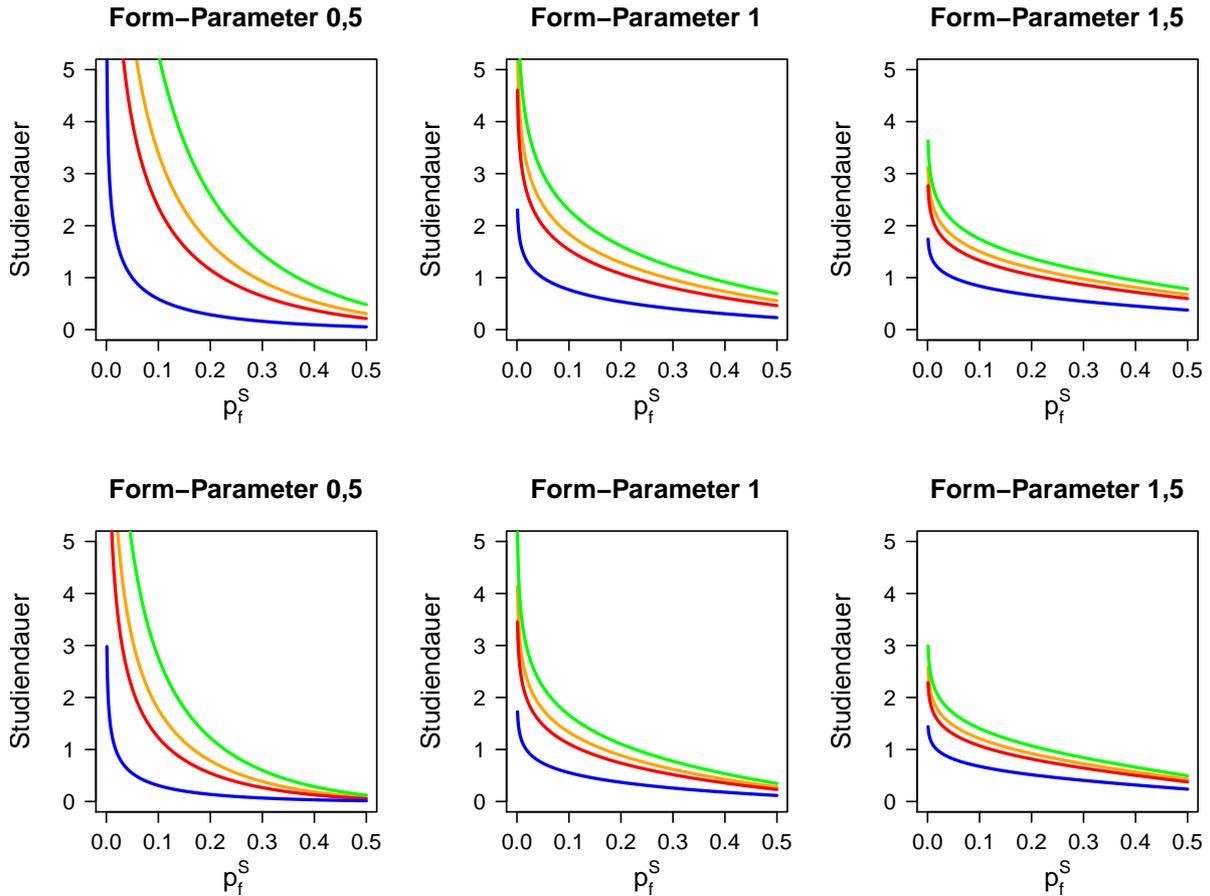


Abbildung 6.1: Abhängigkeit zwischen Studiendauer und Anteil der fest zensierten Patienten aus der Standardgruppe p_f^S nach (6.4), für verschiedene Skalen-Parameter $\lambda \in \{1, 1,25, 1,5, 3\}$ und Anteile von zufällig zensierten Beobachtungen in der Standardgruppe $p_z^S = 10^{-20}\%$ (obere Reihe) und $p_z^S = 25\%$ (untere Reihe).

Setzen wir nun Gleichung (6.4) in Gleichung (6.3) ein, so erhalten wir nach Umformung und Beachtung, dass $p_f^S + p_z^S \neq 1$ (sonst wären alle Personen zensiert)

$$u = \frac{p_z^S \cdot \lambda}{1 - p_z^S - p_f^S}. \quad (6.5)$$

Nach diesen Vorüberlegungen haben wir eine Darstellung des Skalen-Parameters u der Zensierungsverteilung gefunden. Daraus ergeben sich folgende Verteilungen für die Überlebens- bzw. Zensierungszeiten:

$$T_S \sim Wb(\lambda, \vartheta),$$

$$T_N \sim Wb(\lambda, \vartheta),$$

$$T_P \sim Wb(\lambda_P, \vartheta),$$

$$U_i \sim Wb\left(\frac{p_z^S \cdot \lambda}{1 - p_z^S - p_f^S}, \vartheta\right) \quad \text{für } i \in \{1, \dots, n\}.$$

Wir wählen den Anteil an zufälligen und festen Zensierungen basierend auf der Standardgruppe. Nun werden wir diskutieren welche Zensierungsanteile in den anderen beiden Gruppen zu erwarten sind. Da wir annehmen, dass die Verteilungen der Überlebenszeiten in der Standard- und der Novumgruppe gleich sind, erwarten wir in der Novumgruppe denselben Anteil an zensierten Beobachtungen wie in der Standardgruppe. Erwarten wir für Personen aus der Placebogruppe ein geringeres momentanes Ereignisrisiko als für Personen aus der Standard- bzw. Novumgruppe, ist der Anteil an fest zensierten Beobachtungen in der Placebogruppe größer als p_f^S . Nun vergleichen wir die Anteile der zufälligen Zensierungen in den Behandlungsgruppen miteinander. In Abbildung 6.2 sind Dichten der Weibull-Verteilung der Zufallsvariablen $T_S \sim Wb(3; \vartheta)$, $T_P \sim Wb(1; \vartheta)$ und $U_1 \sim Wb(0,33; \vartheta)$ für verschiedene Form-Parameter ϑ abgetragen. Die Wahl eines Skalen-Parameters $u = 0,33$ entspricht einem Anteil von zufällig zensierten Beobachtungen von 10% in der Standardgruppe, wenn der Anteil der fest zensierten Beobachtungen in der Standardgruppe geradezu 0% (hier $10^{-20}\%$ gesetzt) beträgt.

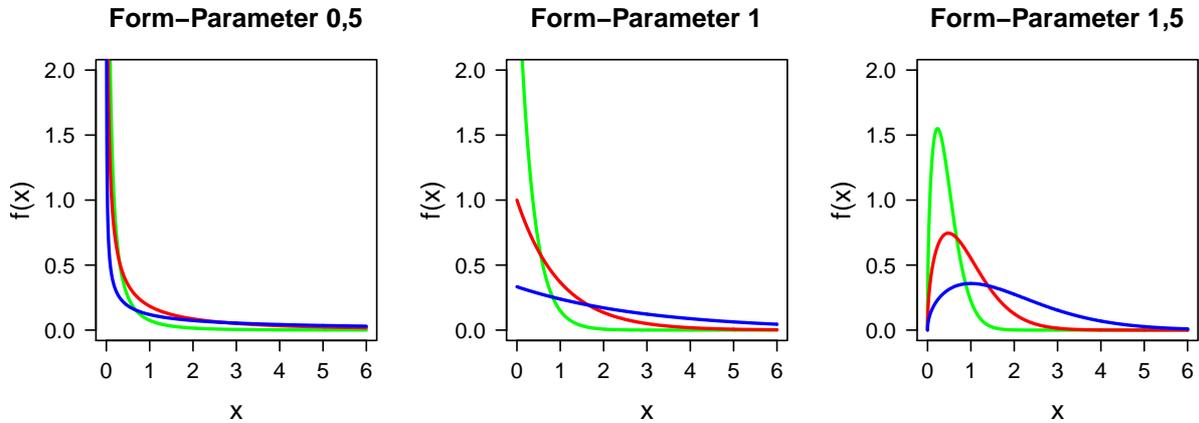


Abbildung 6.2: Dichten der Weibull-Verteilung mit Skalen-Parameter **0,33**, **1** und **3** für verschiedene Form-Parameter.

Die Dichte der Überlebenszeiten von Personen aus der Standard- bzw. Novumgruppe schneidet jeweils die Dichte der Überlebenszeiten von Personen aus der Placebogruppe. Betrachten wir alle Ereignis- und Zensierungszeiten die bis zu diesem Schnittpunkt eingetreten sind, ist der Anteil der zufällig zensierten Personen in der Placebogruppe höher als in der Standard- bzw. Novumgruppe. Werden hingegen die Daten der gesamten Studiendauer betrachtet, ist der Anteil der zensierten Beobachtungen in der Placebogruppe oft höher, da die Dichte der Standard- bzw. Novumgruppe teilweise wesentlich über der der Placebogruppe verläuft. Je mehr zensierte Beobachtungen wir erwarten, desto wesentlich mehr Fallzahlen werden benötigt. Da bei der zufälligen Zensierung die Zeitpunkte der Zensierungen über den gesamten Beobachtungszeitraum verstreut sind und somit kleiner als die Zeitpunkte der fest zensierten Beobachtungen sind, beeinflusst die zufällige Zensierung die Höhe der Fallzahlen stärker als die feste. Mit diesen Ergebnissen können wir Gleichung (6.2) unter Ausnutzung von (6.4) und (6.5) schreiben als

$$L_k = \left[n \cdot w_k \cdot \lambda_k \cdot \frac{1 - p_f^S}{\lambda_k + p_z^S \cdot \lambda / (1 - p_z^S - p_f^S)} \right]. \quad (6.6)$$

Die Anzahl der Ereignisse lässt sich für jede Gruppe nach (6.6) und ihre Lage nach (6.1) bestimmen. Dies kann nun in die Funktion g aus Formel (5.4) eingesetzt werden. Dafür teilen wir jede Summe in drei Teilsummen auf, wobei in jeder Summe über die Ereignisse einer Behandlungsgruppe summiert wird

$$g(\theta, \theta_P, w, L, t) = \frac{g_1(\theta, \theta_P, w, L, t)}{g_2(\theta, \theta_P, w, L, t)} \quad (6.7)$$

mit

$$\begin{aligned}
 g_1(\theta, \theta_P, w, L, t) := & \sum_{k=S,N,P} \sum_{l=1}^{L_k} \left(\frac{f^\theta(t_l^k) w_S [f^\theta(t_l^k) w_N + f_P^{\theta_P}(t_l^k) w_P]}{[f^\theta(t_l) (w_S + w_N) + f_P^{\theta_P}(t_l) w_P]^2} \right. \\
 & - 2\Delta \frac{[f^\theta(t_l^k)]^2 w_S w_N}{[f^\theta(t_l) (w_S + w_N) + f_P^{\theta_P}(t_l) w_P]^2} \\
 & \left. + \Delta^2 \frac{f^\theta(t_l^k) w_N [f^\theta(t_l^k) w_S + f_P^{\theta_P}(t_l^k) w_P]}{[f^\theta(t_l) (w_S + w_N) + f_P^{\theta_P}(t_l) w_P]^2} \right),
 \end{aligned}$$

und

$$\begin{aligned}
 g_2(\theta, \theta_P, w, L, t) := & \left(\sum_{k=S,N,P} \sum_{l=1}^{L_k} \frac{f^\theta(t_l^k) w_S [f^\theta(t_l^k) w_N + f_P^{\theta_P}(t_l^k) w_P]}{[f^\theta(t_l) (w_S + w_N) + f_P^{\theta_P}(t_l) w_P]^2} \right) \\
 & \cdot \left(\sum_{k=S,N,P} \sum_{l=1}^{L_k} \frac{f^\theta(t_l^k) w_N [f^\theta(t_l^k) w_S + f_P^{\theta_P}(t_l^k) w_P]}{[f^\theta(t_l) (w_S + w_N) + f_P^{\theta_P}(t_l) w_P]^2} \right) \\
 & - \left(\sum_{k=S,N,P} \sum_{l=1}^{L_k} \frac{[f^\theta(t_l^k)]^2 w_S w_N}{[f^\theta(t_l) (w_S + w_N) + f_P^{\theta_P}(t_l) w_P]^2} \right)^2.
 \end{aligned}$$

Da sich die Funktion g nicht explizit nach L_k auflösen lässt, kann die erforderliche Anzahl an Ereignissen nicht direkt berechnet werden. Im nächsten Abschnitt entwickeln wir deshalb Algorithmen, mit denen der benötigte Stichprobenumfang pro Gruppe berechnet werden kann.

6.2 Algorithmen

6.2.1 Berechnung der Fallzahlen bei festen Randomisierungsverhältnissen in die Gruppen

Zunächst sind wir daran interessiert wie groß der gesamte Stichprobenumfang gewählt werden muss, wenn ein festes Verhältnis der Personen pro Gruppe zueinander gegeben ist.

Algorithmus zur Bestimmung der Fallzahlen insgesamt bei festen Randomisierungsverhältnissen in die Gruppen

Input: $\theta = (\lambda, \vartheta)$, $\theta_P = (\lambda_P, \vartheta)$, $w_S, w_N, w_P, p_f^S, p_z^S, \alpha, \gamma, \Delta$

Toleranzwert: $\epsilon > 0$

Startwert: n

1) berechne $L_k = \left\lceil n \cdot w_k \cdot \lambda_k \cdot \frac{1-p_f^S}{\lambda_k + (p_z^S \cdot \lambda) / (1-p_z^S - p_f^S)} \right\rceil$ für $k = S, N, P$ (nach (6.6))
und $L = L_S + L_N + L_P$

2) bestimme die Lage der Ereigniszeitpunkte $t_1^k, \dots, t_{L_k}^k$ für $k = S, N, P$ (nach (6.1))

3) bestimme $g(\theta, \theta_P, w, L, t)$ (nach (6.7))

4) solange $\left| g(\theta, \theta_P, w, L, t) - \left(\frac{\beta_N - \Delta \beta_S}{u_{1-\gamma} - u_{1-\alpha}} \right)^2 \right| > \epsilon$

- verringere n falls $g(\theta, \theta_P, w, L, t) - \left(\frac{\beta_N - \Delta \beta_S}{u_{1-\gamma} - u_{1-\alpha}} \right)^2 > \epsilon$
- erhöhe n falls $-g(\theta, \theta_P, w, L, t) + \left(\frac{\beta_N - \Delta \beta_S}{u_{1-\gamma} - u_{1-\alpha}} \right)^2 > \epsilon$

gehe zu Schritt 1)

5) falls $\left| g(\theta, \theta_P, w, L, t) - \left(\frac{\beta_N - \Delta \beta_S}{u_{1-\gamma} - u_{1-\alpha}} \right)^2 \right| < \epsilon$

Abbruch: die erforderliche gesamte Personenanzahl n ist gefunden.

Output: insgesamt erforderlicher Stichprobenumfang n_{fest}^* und Studiendauer b

Im nächsten Abschnitt geben wir an, wie wir einen geeigneten Startwert für den insgesamten Stichprobenumfang n wählen können.

Startwert für den Algorithmus

Die Wahl des Startwertes für den insgesamten Stichprobenumfang n beeinflusst die Rechenzeit des Algorithmus. Um einen geeigneten Startwert zu finden, interessiert uns zunächst nicht, zu welchen Zeitpunkten im Studienverlauf genau die Ereignisse eintreten sondern nur, ob sie eintreten. In diesem Fall spricht man von binären Endpunkten (parametrische Verfahren für binäre Endpunkte wurden in der Literatur von [Kieser and Friede \[2007\]](#) oder auch [Mielke and Munk](#) betrachtet). Vereinfachend nehmen wir an, dass alle Ereigniszeitpunkte an demselben Zeitpunkt eintreten, d.h. $t_l = t$ für $l \in \{1, \dots, L\}$. Wir werden zeigen, dass sich die Funktion g aus Gleichung (5.4) in diesem Fall zu $4/L$ vereinfacht und geben schließlich eine Formel an, nach der die optimale Personenanzahl insgesamt und deren optimale Aufteilung berechnet werden kann. Gleichzeitig erhalten wir so einen Einblick in die Gruppenverhältnisse zueinander.

Die L Summanden in den Summen aus Funktion g in (5.4) hängen nun nicht mehr vom Summationsindex ab und sind somit alle identisch. Aus Gleichung (5.2)-(5.4) ergibt sich:

$$L \approx \frac{(w_N + \kappa w_P) w_S - 2 \Delta w_S w_N + \Delta^2 w_N (\kappa w_P + w_S)}{\underbrace{(w_N + \kappa w_P + w_S)^2 \left(-\frac{w_N^4 w_S^4}{(w_N + \kappa w_P + w_S)^8} + \frac{w_N (w_N + \kappa w_P) w_S (\kappa w_P + w_S)}{(w_N + \kappa w_P + w_S)^4} \right)}_{:=\tilde{g}(w_S, w_N)}} \cdot \left(\frac{u_{1-\gamma} - u_{1-\alpha}}{\beta_N - \Delta \cdot \beta_S} \right)^2, \quad (6.8)$$

wobei $f_P(t)/f(t) =: \kappa$ und $w_P = 1 - w_S - w_N$.

Wir wollen nun w_S und w_N gerade so wählen, dass wir für die gegebenen Parameterwerte die minimal erforderliche Anzahl an Ereignissen L benötigen. Dafür untersuchen wir die Funktion

$$\tilde{g} : W \rightarrow \mathbb{R}, \text{ mit } W := \{(w_S, w_N) \in (0, 1)^2 : w_S + w_N \in (0, 1)\}$$

auf Minima.

Satz 6.1

Angenommen alle Ereignisse treten zum selben Zeitpunkt ein, also $t_l = t$ für $l \in \{1, \dots, L\}$ gilt, dann sind die folgenden Anzahlen an Personen pro Patientengruppe in folgendem Sinn optimal: Sie stellen die kleinstmöglichen Stichprobenumfänge dar, die zu einer vorgegebenen Güte von γ führen, bei einem Niveau von α , einer Nichtunterlegenheitsgrenze Δ , Hazard Quotienten β_S bzw. β_N unter der Alternative $\beta_S = \beta_N \neq 0$:

$$\begin{aligned} n_S^* &= 4 \cdot \left(\frac{u_{1-\gamma} - u_{1-\alpha}}{\beta_N - \Delta \cdot \beta_S} \right)^2 \cdot \frac{1}{\psi^*} \cdot \frac{\Delta f_P(t)/f(t)}{f_P(t)/f(t)(1 + \Delta) + 1 - \Delta} \\ n_N^* &= 4 \cdot \left(\frac{u_{1-\gamma} - u_{1-\alpha}}{\beta_N - \Delta \cdot \beta_S} \right)^2 \cdot \frac{1}{\psi^*} \cdot \frac{f_P(t)/f(t)}{f_P(t)/f(t)(1 + \Delta) + 1 - \Delta} \\ n_P^* &= 4 \cdot \left(\frac{u_{1-\gamma} - u_{1-\alpha}}{\beta_N - \Delta \cdot \beta_S} \right)^2 \cdot \frac{1}{\psi^*} \cdot \frac{1 - \Delta}{f_P(t)/f(t)(1 + \Delta) + 1 - \Delta}, \end{aligned}$$

mit

$$\psi^* = P(T_S < b, T_S < U_1) w_S^* + P(T_N < b, T_N < U_1) w_N^* + P(T_P < b, T_P < U_1) w_P^*,$$

wobei

$$\begin{aligned} w_S^* &= \Delta w_N^* \\ w_N^* &= \frac{\kappa}{\kappa(1 + \Delta) + 1 - \Delta} \\ w_P^* &= 1 - w_S^* - w_N^*, \end{aligned}$$

f die Dichte der Überlebenszeiten von Personen aus der Standard- und Novumgruppe und f_P Dichte der Überlebenszeiten von Personen aus der Placebogruppe, $\kappa = f_P(t)/f(t)$ und $u_{1-\alpha}$ das $(1 - \alpha)$ -Quantil der Standardnormalverteilung bezeichnet. Insgesamt beträgt der benötigte Stichprobenumfang dann:

$$n^* = n_S^* + n_N^* + n_P^* = 4 \cdot \left(\frac{u_{1-\gamma} - u_{1-\alpha}}{\beta_N - \Delta \cdot \beta_S} \right)^2 \cdot \frac{1}{\psi^*}. \quad (6.9)$$

Beweis: Um den Satz zu beweisen, gehen wir folgendermaßen vor:

- 1) Zunächst zeigen wir, dass $\tilde{g} : W \rightarrow \mathbb{R}$ strikt konvex ist.
- 2) Dann minimieren wir $\tilde{g} : W \rightarrow \mathbb{R}$.
 - a) Als erstes ermitteln wir alle Extrema von \tilde{g} und werden feststellen, dass ein Extremum in W liegt.
 - b) Als zweites zeigen wir, dass es sich hierbei tatsächlich um ein Minimum handelt.

zu 1) Um zu zeigen, dass \tilde{g} strikt konvex ist, benutzen wir die Aussage des folgenden Satzes:

Satz 6.2

Eine Funktion $h : \mathbb{R}^d \rightarrow \mathbb{R}$, $h \in \mathcal{C}^2$ ist genau dann strikt konvex, wenn $Hess(h)$ positiv definit ist, d.h. $x^T Hess(h)x > 0$ für alle $x \in \mathbb{R}^d$, wobei $x \neq 0$. \square

Damit ergibt sich in unserem Fall für $(w_S, w_N) \in W$:

$$Hess(\tilde{g}) = \begin{pmatrix} \frac{\partial^2 \tilde{g}}{\partial w_S^2} & \frac{\partial^2 \tilde{g}}{\partial w_S \partial w_N} \\ \frac{\partial^2 \tilde{g}}{\partial w_N \partial w_S} & \frac{\partial^2 \tilde{g}}{\partial w_N^2} \end{pmatrix} =: \begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix}$$

und somit

$$\begin{aligned} & (w_S, w_N)^T Hess(\tilde{g})(w_S, w_N) \\ &= w_S^2 a_1 + w_N^2 a_4 + w_S w_N (a_3 + a_2) \\ &= 2 \frac{(1 - \Delta)^2 w_S w_N (w_S + w_N)^2 + (1 - w_S - w_N)^3 \kappa^2 (w_S + \Delta w_N)}{\kappa w_S w_N (1 - w_S - w_N)^3} \\ &> 0. \end{aligned}$$

zu 2a) Zunächst bestimmen wir die partiellen Ableitungen von \tilde{g} :

$$\begin{aligned} \frac{\partial \tilde{g}}{\partial w_S} &= \frac{-2\Delta w_N w_S^2 + w_S^2 (w_N + \kappa (w_N + w_S - 1)^2 - \kappa^2 (w_N + w_S - 1)^2)}{\kappa w_N w_S^2 (-1 + w_S + w_N)^2} \\ &\quad + \frac{\Delta^2 w_N (w_S^2 + \kappa^2 (w_N - 1) (w_N + w_S - 1)^2 - \kappa w_N (w_N + w_S - 1)^2)}{\kappa w_N w_S^2 (-1 + w_S + w_N)^2} \\ \frac{\partial \tilde{g}}{\partial w_N} &= \frac{-2\Delta w_N^2 w_S + \Delta^2 w_N^2 (w_S + \kappa (w_S + w_N - 1)^2 - \kappa^2 (w_S + w_N - 1)^2)}{\kappa w_N w_S^2 (-1 + w_S + w_N)^2} \\ &\quad + \frac{w_S (w_N^2 + \kappa^2 (w_S - 1) (w_S + w_N - 1)^2 - \kappa w_S (w_S + w_N - 1)^2)}{\kappa w_N w_S^2 (-1 + w_S + w_N)^2}. \end{aligned}$$

Nun ermitteln wir alle Paare (w_S, w_N) , bei denen beide Ableitungen gleich sind, d.h.:

$$\frac{\partial \tilde{g}}{\partial w_S} - \frac{\partial \tilde{g}}{\partial w_N} = \frac{(\Delta w_N - w_S) (\Delta w_N + w_S) (-\kappa - w_N + \kappa w_N - w_S + \kappa w_S)}{w_N^2 w_S^2} = 0.$$

Dies ist der Fall, wenn

(i) $\Delta w_N = w_S$ oder

(ii) $-\Delta w_N = w_S$ oder

(iii) $-\kappa - w_N + \kappa w_N - w_S + \kappa w_S = 0 \Leftrightarrow w_S = \frac{\kappa}{\kappa-1} - w_N$.

Jetzt setzen wir diese Lösungen in die Ableitungen ein, bestimmen w_S gerade so, dass beide Ableitungen 0 sind und überprüfen, welche der Lösungen in W liegt.

(i) $\Delta w_N = w_S$

Lösung 1: $\left(w_S = \Delta w_N, w_N = \frac{\kappa}{\kappa(1+\Delta)+1-\Delta}\right) \in W$, da
 $w_N > 0, w_S > 0, w_S + w_N = \frac{\kappa(1+\Delta)}{\kappa(1+\Delta)+1-\Delta} > 0$ und $w_S + w_N < 1$.

Lösung 2: $\left(w_S = \Delta w_N, w_N = \frac{\kappa}{\kappa(1+\Delta)-1+\Delta}\right) \notin W$, da
 $w_S + w_N = \frac{1+\Delta}{\frac{\Delta-1}{\kappa}+1+\Delta} < 1$ nur falls $\kappa < 0$, aber κ ist ein Quotient zweier Dichten und somit positiv.

(ii) $\Delta w_N = -w_S$

Lösung 3: $\left(w_S = -\Delta w_N, w_N = -\frac{\kappa}{(-1+\Delta)(-1+\kappa)}\right) \notin W$

Lösung 4: $\left(w_S = -\Delta w_N, w_N = \frac{\kappa}{(-1+\Delta)(1+\kappa)}\right) \notin W$

(iii) $w_S = \frac{\kappa}{\kappa-1} - w_N$

Lösung 5: $\left(w_S = \frac{\kappa}{\kappa-1} - w_N, w_N = -\frac{\kappa}{(-1+\Delta)(-1+\kappa)}\right) \notin W$, da

$$w_N = \frac{\kappa}{(1-\Delta)(\kappa-1)} > 0, \text{ nur falls } \kappa > 1$$

$w_S = \frac{\kappa\Delta}{(\Delta-1)(\kappa-1)} > 0$, nur falls $\kappa < 1$. Es kann also kein gemeinsames κ gefunden werden so dass sowohl $w_N > 0$ als auch $w_S > 0$ gilt.

Die einzige Lösung, die in W liegt, ist Lösung 1, also $w^* = \left(\frac{\Delta\kappa}{\kappa(1+\Delta)+1-\Delta}, \frac{\kappa}{\kappa(1+\Delta)+1-\Delta}\right)$.

zu 2b) Da $Hess(\tilde{g})$ in ganz W positiv definit ist und $w^* \in W$, ist $Hess(\tilde{g})$ an der Stelle w^* positiv definit. Auf dem Rand von W kann es keine Punkte mit kleinerem Funktionswert geben.

Somit haben wir mit w^* das globale Minimum von \tilde{g} in W gefunden. Die optimalen Gruppeneinteilungen lauten somit:

$$\begin{aligned} w_S^* &= \Delta w_N^* \\ w_N^* &= \frac{\kappa}{\kappa(1+\Delta) + 1 - \Delta} \\ w_P^* &= 1 - w_S^* - w_N^*. \end{aligned}$$

Für sie gilt das Verhältnis

$$w_S^* : w_N^* : w_P^* = \Delta : 1 : \frac{f(t)}{f_P(t)} (1 - \Delta).$$

Nun bestimmen wir die Anzahl der benötigten Ereignisse. Da die Funktion \tilde{g} für die optimale Gruppeneinteilung den Wert 4 annimmt, erhalten wir für Gleichung (6.8)

$$L^* \approx 4 \cdot \left(\frac{u_{1-\gamma} - u_{1-\alpha}}{\beta_N - \Delta \cdot \beta_S} \right)^2.$$

Unter der Annahme, dass es keine zensierten Beobachtungen geben wird, entspricht dies dem insgesamt benötigten Stichprobenumfang n^* .

Jetzt berücksichtigen wir zusätzlich noch zensierte Personen wie in Kapitel 5.2 beschrieben und erhalten (da $L^* = n_k^*/w_k^* \cdot \psi^*$ gilt) für die optimale Personenanzahl pro Gruppe:

$$\begin{aligned} n_S^* &= 4 \cdot \left(\frac{u_{1-\gamma} - u_{1-\alpha}}{\beta_N - \Delta \cdot \beta_S} \right)^2 \cdot \frac{1}{\psi^*} \cdot \frac{\Delta \kappa}{\kappa(1+\Delta) + 1 - \Delta} \\ n_N^* &= 4 \cdot \left(\frac{u_{1-\gamma} - u_{1-\alpha}}{\beta_N - \Delta \cdot \beta_S} \right)^2 \cdot \frac{1}{\psi^*} \cdot \frac{\kappa}{\kappa(1+\Delta) + 1 - \Delta} \\ n_P^* &= 4 \cdot \left(\frac{u_{1-\gamma} - u_{1-\alpha}}{\beta_N - \Delta \cdot \beta_S} \right)^2 \cdot \frac{1}{\psi^*} \cdot \frac{1 - \Delta}{\kappa(1+\Delta) + 1 - \Delta} \end{aligned}$$

mit

$$\psi^* = P(T_S < b, T_S < U_1) w_S^* + P(T_N < b, T_N < U_1) w_N^* + P(T_P < b, T_P < U_1) w_P^*.$$

Insgesamt beträgt der benötigte Stichprobenumfang dann:

$$n^* = n_S^* + n_N^* + n_P^* = 4 \cdot \left(\frac{u_{1-\gamma} - u_{1-\alpha}}{\beta_N - \Delta \cdot \beta_S} \right)^2 \cdot \frac{1}{\psi^*}. \quad \square$$

Aus diesen Formeln lässt sich nun leicht der erforderliche Stichprobenumfang pro Gruppe berechnen, der als Startwert für den insgesamten Stichprobenumfang n für unseren Algorithmus benutzt werden kann.

Ergebnisse des Algorithmus zur Bestimmung der Fallzahlen insgesamt bei festen Randomisierungsverhältnissen in die Gruppen

Den zu Beginn dieses Abschnittes vorgestellten Algorithmus wollen wir nun anwenden: Üblicherweise werden in der Literatur Aufteilungen der Personen im Verhältnis $n_S:n_N:n_P$ von 1:1:1, 2:2:1, 1:2:1 und 2:3:1 betrachtet. Zusätzlich betrachten wir auch wieder von der Nichtunterlegenheitsgrenze Δ abhängige Aufteilungen $\Delta:1:(1-\Delta)$ für $\Delta \in \{0,5; 0,65; 0,8\}$. Für $\Delta = 0,8$ ergibt sich dann die Aufteilung 4:5:1, für $\Delta = 0,65$ ist es 13:20:7 und bei $\Delta = 0,5$ erhalten wir 1:2:1. Da die Aufteilung 2:3:1 sehr ähnlich zu 13:20:7 ist, beschränken wir uns aus Gründen der Übersichtlichkeit auf die Aufteilung 2:3:1. Für verschiedene Parameterwahlen für $HQ \in [1,25; 3]$ und $\vartheta \in \{0,5; 1; 1,5\}$ haben wir unter der Alternative $\beta_S = \beta_N$ basierend auf Weibull-verteilten Überlebens- und Zensierungszeiten für die oben angegebenen Personenverhältnisse die benötigten gesamten Fallzahlen bezüglich eines nominalen Niveaus von $\alpha = 2,5\%$ bzw. $\alpha = 5\%$ berechnet um eine gewünschte Güte von $\gamma = 80\%$ zu erreichen. Den Parameter λ_P setzen wir auf $\lambda_P = 1$. Wir berücksichtigen keine zensierten Beobachtungen. Wie die Studiendauer b bei den verschiedenen Parameterkonstellationen zu wählen ist, ist Abbildung 6.1 zu entnehmen. Als Toleranzwert haben wir $\epsilon = \left(\frac{\beta_N - \Delta\beta_S}{(u_{1-\gamma} - u_{1-\alpha})}\right)^2 / 100$ gewählt. Die wesentlichen Ergebnisse sind in Abbildung 6.3 dargestellt. Da die Wahl des Form-Parameters der Weibull-Verteilungen ϑ die Höhe der Fallzahlen kaum beeinflusst, haben wir die Berechnungen nur für $\vartheta = 0,5$ angegeben.

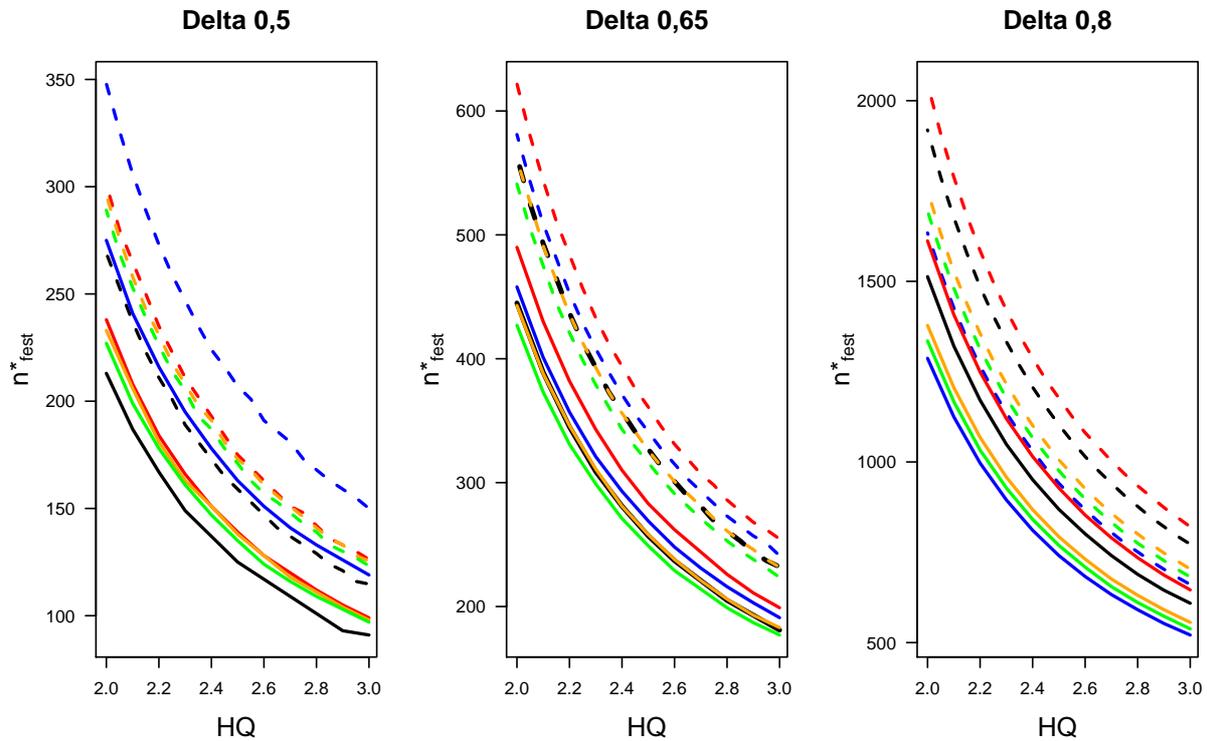


Abbildung 6.3: Fallzahlen n_{fest}^* für verschiedene Aufteilungen der Personen auf die Gruppen nach $n_S:n_N:n_P$ von **1:1:1**, **1:2:1**, **2:2:1**, **2:3:1** und **4:5:1** in Abhängigkeit des Hazard Quotienten (HQ). Die Berechnungen erfolgten unter der Alternative $\beta_S = \beta_N$ bezüglich $\vartheta = 0,5$, eines Niveaus von 2,5% (---) bzw. 5% (—) und einer gewünschten Güte von 80%.

Es wird deutlich, dass die Anzahl der benötigten Personen von Δ abhängt. Bei den betrachteten Personenverhältnissen fällt auf, dass die gesamte Anzahl an Personen für $\Delta = 0,5$ bei dem Verhältnis 1:2:1 am geringsten ist, für $\Delta = 0,65$ bei der Aufteilung 2:3:1 und für $\Delta = 0,8$ bei 4:5:1.

6.2.2 Berechnung der Fallzahlen mit optimaler Aufteilung in die Gruppen

Nun wollen wir untersuchen, ob die Anzahl der insgesamt benötigten Personen durch optimale Aufteilung der Personen in die Behandlungsgruppen noch verringert werden kann. Aus (5.1) wird deutlich, dass die asymptotische Güte um so größer ist, je kleiner folgender

Ausdruck ist

$$\widetilde{\text{Var}(\widehat{\beta}_N)} - 2 \cdot \Delta \cdot \widetilde{\text{Cov}(\widehat{\beta}_S, \widehat{\beta}_N)} + \Delta^2 \cdot \widetilde{\text{Var}(\widehat{\beta}_S)} \stackrel{(5.3)}{=} g(\theta, \theta_P, w, L, t).$$

Wir wollen als nächstes die Aufteilung ermitteln, bei der wir die größte Güte erwarten. Dazu ist die Funktion g aus (6.7) in $w = (w_S, w_N, w_P)$ zu minimieren (bei fester Wahl der anderen Parameter) unter den Nebenbedingungen $w_S + w_N + w_P = 1$ und $w_S, w_N, w_P \in (0, 1)$.

Aufgrund der komplexen Struktur der Funktion g sowie der vorhandenen Nebenbedingungen, kann das gesuchte Minimum nicht analytisch bestimmt werden. Wir verwenden daher einen numerischen Algorithmus, der das gesuchte Minimum unter Beachtung der Nebenbedingungen findet. Der Nelder-Mead Algorithmus ist ein Minimierungs-Verfahren für nichtlineare und nicht notwendig differenzierbare von mehreren Parametern abhängige Funktionen ohne Nebenbedingungen. Er ist ein sehr häufig verwendetes Verfahren. Die Nebenbedingungen der Funktion g berücksichtigen wir, indem wir $w_P = 1 - w_S - w_N$ setzen und den Funktionswert von g bei verletzten Nebenbedingungen auf einen sehr hohen Wert setzen (d.h. für $w_S \notin (0, 1)$ und/oder $w_N \notin (0, 1)$ und/oder $w_S + w_N \notin (0, 1)$). Für gegebene θ, θ_P, L, t definieren wir $g_{\theta, \theta_P, L, t} : \mathbb{R}^2 \rightarrow \mathbb{R}$ durch

$$g_{\theta, \theta_P, L, t}(w_S, w_N) := \begin{cases} g(\theta, \theta_P, (w_S, w_N, 1 - w_S - w_N), L, t) & : \text{für } w_S, w_N, w_S + w_N \in (0, 1) \\ 10^{60} & : \text{sonst.} \end{cases}$$

Diese auf ganz \mathbb{R}^2 definierte Funktion ist nicht differenzierbar.

Wie der Nelder-Mead Algorithmus funktioniert, ist in [Conn et al. \[2009\]](#) beschrieben und wird nun kurz für unser Minimierungsproblem skizziert.

Die Idee des Verfahrens besteht darin, ein Dreieck, dessen Eckpunkte nicht auf einer Geraden liegen (hierbei handelt es sich um einen speziellen Simplex), zu erzeugen. In einem Iterationsschritt wird für jeden Eckpunkt der Wert der Zielfunktion $g_{\theta, \theta_P, L, t}$ berechnet. Unter den 3 Punkten wird der beste und der schlechteste ermittelt. Der schlechteste wird dann entweder durch einen neuen ersetzt (so dass das Dreieck bezüglich des neuen Punktes um eine Seite kippt, sich streckt oder zusammenzieht) oder das Dreieck zieht sich um den bisher besten Punkt zusammen. So verlagern sich die Simplexe schließlich in Richtung des Minimums und ziehen sich um dieses zusammen.

Anhand vieler Parameterkonstellationen haben wir den Graphen der Funktion (der es nahe legt, dass es sich bei der Funktion g um eine konvexe handelt) betrachtet und mit dem vorgestellten Verfahren jeweils das gesuchte Minimum gefunden.

Algorithmus zur Bestimmung der Fallzahlen mit optimaler Aufteilung in die Gruppen

Input: $\theta = (\lambda, \vartheta)$, $\theta_P = (\lambda_P, \vartheta)$, $p_f^S, p_z^S, \alpha, \gamma, \Delta$

Toleranzwert: $\epsilon > 0$

Startwert: n, w_S, w_N, w_P

1) bestimme die Anzahl der Ereignisse in Abhängigkeit von w_k , also

$$L_k(w_k) = \left\lceil n \cdot w_k \cdot \lambda_k \cdot \frac{1-p_f^S}{\lambda_k + p_z^S \cdot \lambda / (1-p_z^S - p_f^S)} \right\rceil \text{ für } k = S, N, P \text{ (nach (6.6))}$$

2) bestimme die Lage der Ereigniszeitpunkte $t_1^k, \dots, t_{L_k(w_k)}^k$ für $k = S, N, P$ (nach 6.1)

3) bestimme $g(\theta, \theta_P, w, L(w), t)$ (nach (6.7)) und ermittle mit der Simplexmethode nach Nelder-Mead (wie auf Seite 70 beschrieben) die Aufteilung $w^* = (w_S^*, w_N^*, w_P^*)$, für die g minimal ist

4) solange $\left| g(\theta, \theta_P, w^*, L(w^*), t) - \left(\frac{\beta_N - \Delta \beta_S}{u_{1-\gamma} - u_{1-\alpha}} \right)^2 \right| > \epsilon$

- verringere n falls $g(\theta, \theta_P, w^*, L(w^*), t) - \left(\frac{\beta_N - \Delta \beta_S}{u_{1-\gamma} - u_{1-\alpha}} \right)^2 > \epsilon$
- erhöhe n falls $-g(\theta, \theta_P, w^*, L(w^*), t) + \left(\frac{\beta_N - \Delta \beta_S}{u_{1-\gamma} - u_{1-\alpha}} \right)^2 > \epsilon$

gehe zu Schritt 1)

5) falls $\left| g(\theta, \theta_P, w^*, L(w^*), t) - \left(\frac{\beta_N - \Delta \beta_S}{u_{1-\gamma} - u_{1-\alpha}} \right)^2 \right| < \epsilon$

Abbruch, die erforderliche gesamte Personenanzahl n^* und deren beste Aufteilung in Gruppen $w^* = (w_S^*, w_N^*, w_P^*)$ ist gefunden

Output: insgesamt erforderlicher Stichprobenumfang n_{semip}^* ,
dessen Aufteilung in die Gruppen (w_S^*, w_N^*, w_P^*) und
Studiendauer b

Der Startwert für n kann z.B. aus Formel (6.9) bestimmt werden. Als Startwert für w_S, w_N, w_P kann beispielsweise $w_N = 1/2, w_S = \Delta w_N = \Delta/2$ und $w_P = 1 - w_N - w_S = (1-\Delta)/2$ verwendet werden.

Ergebnisse des Algorithmus zur Bestimmung der Fallzahlen mit optimaler Aufteilung in die Gruppen

Wir werden nun die erforderlichen Fallzahlen n_{fest}^* für die üblicherweise verwendeten festen Aufteilungen mit n_{semip}^* aus dem eben vorgestellten Algorithmus vergleichen. Dazu berechnen wir den benötigten Stichprobenumfang für die Nichtunterlegenheitsgrenzen $\Delta \in \{0,5; 0,65; 0,8\}$, die Form-Parameter $\vartheta \in \{0,5; 1; 1,5\}$ und die Hazard Quotienten $HQ \in [1,25; 3]$. Den Skalen-Parameter λ_P setzen wir auf $\lambda_P = 1$ den Toleranzwert ϵ auf $\epsilon = ((\beta_N - \Delta\beta_S)/(u_{1-\gamma} - u_{1-\alpha}))^2/100$, für das Niveau erlauben wir $\alpha = 2,5\%$ und für die gewünschte Güte $\gamma = 80\%$ unter der Alternative $\beta_S = \beta_N$. Unsere Berechnungen führen wir einmal mit dem Verfahren aus Abschnitt 6.2.1 und außerdem mit dem Verfahren aus Abschnitt 6.2.2 für die festen Aufteilungen 1:1:1, 2:2:1, 2:3:1, 3:3:1 und $\Delta:1:(1 - \Delta)$ durch.

In Abbildungen 6.4 und 6.5 ist die prozentuale Abweichung der Fallzahlen n_{fest}^* bei fester Aufteilung von den Fallzahlen n_{semip}^* bei optimaler Aufteilung dargestellt. In anderen Worten: auf der y-Achse ist $(n_{\text{fest}}^*/n_{\text{semip}}^* - 1) \cdot 100\%$ abgetragen. Hier wurden keine zensierten Beobachtungen berücksichtigt, da wir für verschiedene Grade an zufällig und fest zensierten Beobachtungen dieselben Aussagen erhalten. Da verschiedene Wahlen für den Form-Parameter ϑ kaum Einfluss auf die Höhe der Fallzahlen haben, betrachten wir nur $\vartheta = 0,5$.

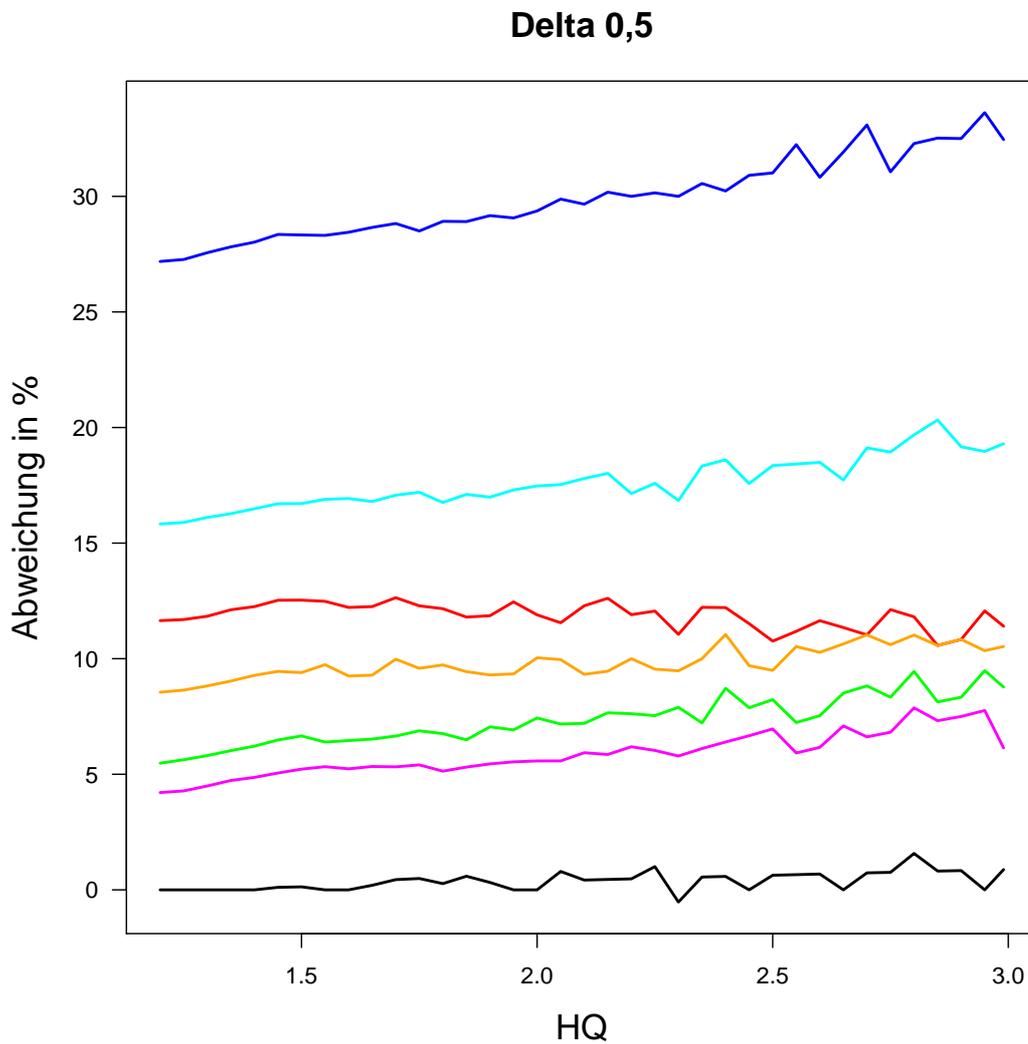


Abbildung 6.4: Prozentsatz, um den die Fallzahlen n_{fest}^* bezüglich n_{semip}^* für verschiedene Hazard Quotienten (HQ) anwachsen. Als feste Aufteilungen für $n_S:n_N:n_P$ betrachten wir **1:1:1**, **1:2:1**, **2:2:1**, **2:3:1**, **3:3:1**, **4:5:1** und **13:20:7**. Für Parameterwahlen $\vartheta = 0,5$, ein Niveau von 2,5%, einer Güte von 80% unter der Alternative $\beta_S = \beta_N$.

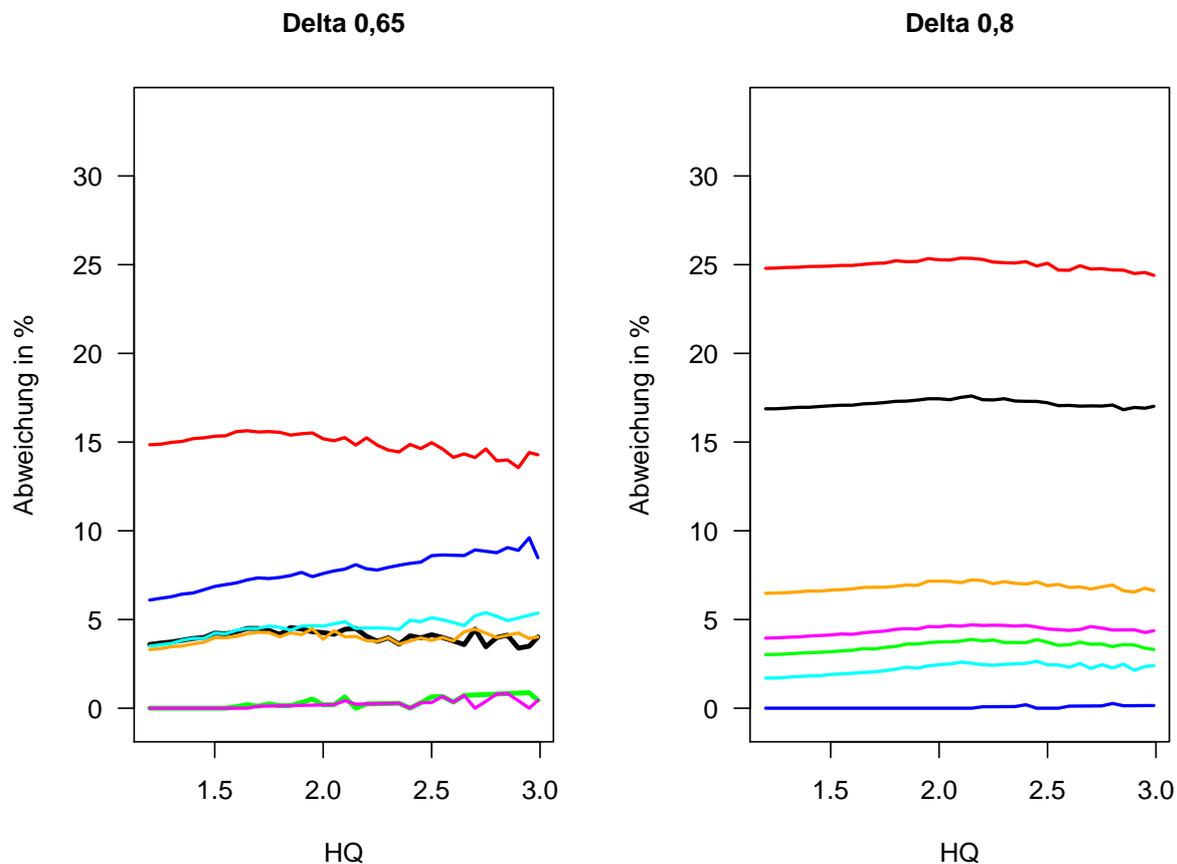


Abbildung 6.5: Prozentsatz, um den die Fallzahlen n_{fest}^* bezüglich n_{semip}^* für verschiedene Hazard Quotienten (HQ) anwachsen. Als feste Aufteilungen für $n_S:n_N:n_P$ betrachten wir **1:1:1**, **1:2:1**, **2:2:1**, **2:3:1**, **3:3:1**, **4:5:1** und **13:20:7**. Für Parameterwahlen $\vartheta = 0,5$, ein Niveau von 2,5%, einer Güte von 80% unter der Alternative $\beta_S = \beta_N$.

In der Tat hat die Aufteilung der Patienten auf die Gruppen eine erhebliche Auswirkung auf die Gesamtfallzahl. Auch sieht man, dass es keine feste Aufteilung gibt, die in allen dargestellten Fällen optimal ist, so dass grundsätzlich der variablen Aufteilung aus Abschnitt 6.2.2 der Vorzug zu geben ist. Hier bestätigt sich unsere Intuition aus Kapitel 4.2, denn die Kurven zu dem festen Personenverhältnis $n_S:n_N:n_P = \Delta:1:(1-\Delta)$ für das jeweils betrachtete Δ verlaufen nahe der Nulllinie und kommen somit dem benötigten Stichprobenumfang bei Verwendung unseres Verfahrens sehr nahe.

Depressionsstudie (Fortsetzung)

In unserer Depressionsstudie sind die Personen im Verhältnis $n_S:n_N:n_P=2:2:1$ aufgeteilt. Wird solch eine Aufteilung der Personen auf die Gruppen gewählt, weichen die Fallzahlen bei den eben betrachteten Parameterwahlen um etwa 4% bis 11% von unserem Verfahren aus Abschnitt 6.2.2 nach oben ab, je nachdem wie die Nichtunterlegenheitsgrenze Δ gewählt wird (vgl. Abbildungen 6.4 und 6.5).

6.3 Untersuchung der Testverfahren mittels Simulationen

Simulationen der Güte

Nun untersuchen wir für die mit unserem Algorithmus aus Abschnitt 6.2.2 berechneten Fallzahlen die Güte für die in Kapitel 4 vorgestellten Tests. Wir haben die Stichprobenumfänge n_{semip}^* unter der Alternative $\beta_S = \beta_N$ für verschiedene Hazard Quotienten $HQ \in \{1,25; 3\}$, Form-Parameter $\vartheta \in \{0,5; 1; 1,5\}$ und Nichtunterlegenheitsgrenzen $\Delta \in \{0,5; 0,65; 0,8\}$ berechnet. Dabei setzen wir den Skalen-Parameter λ_P ohne Einschränkung auf $\lambda_P = 1$, erlauben ein Niveau von $\alpha = 2,5\%$ und wählen eine gewünschte Güte von $\gamma = 80\%$. Als Toleranzwert im Algorithmus haben wir $\epsilon = ((\beta_N - \Delta\beta_S)/(u_{1-\gamma} - u_{1-\alpha}))^2/100$ benutzt. Wir führen unsere Simulationen und Berechnungen sowohl ohne als auch mit Berücksichtigung von zensierten Beobachtungen durch. Für die Simulationen bei zensierten Beobachtungen wählen wir den Parameter der Zensierungsverteilung und die Studiendauer gerade so, dass $p_f^S = p_z^S = 25\%$, d.h. wir erwarten einen Anteil von 25% an fest und einen Anteil von 25% an zufällig zensierten Beobachtungen in der Standardgruppe. Die Stichprobenumfänge n_{semip}^* teilen wir auch im Verhältnis 1:1:1, 2:2:1 und $\Delta:1:(1-\Delta)$ auf und simulieren die Güte. Dadurch können wir untersuchen, wie viel an Güte bei diesen festen Aufteilungen im Gegensatz zu der variablen verloren geht. Insgesamt haben wir jeweils 10.000 Wiederholungen durchgeführt. In Tabelle 6.1 sind die wesentlichen Ergebnisse dargestellt.

$p_f^S = p_z^S$	Δ	HQ	n_{semip}^*	$n_S:n_N:n_P$	$H_0^{Ueb,S}$	$H_0^{Ueb,N}$	H_0^{Nu}	$H_0^{Ueb,S} \cup H_0^{Nu}$	$H_0^{Ueb,N} \cup H_0^{Nu}$		
10 ⁻²⁰ %	0,5	1,25	2.523	optimal	97,60	99,56	79,41	77,41	79,41		
				1:1:1	99,38	99,40	74,54	74,05	74,54		
				2:2:1	98,33	98,07	75,73	74,79	75,71		
				$\Delta:1:(1-\Delta)$	97,72	99,51	79,40	77,58	79,40		
		3	113	optimal	97,87	99,37	79,99	78,43	79,99		
				1:1:1	99,51	99,75	76,68	76,32	76,68		
				2:2:1	97,96	98,01	76,86	75,87	76,85		
				$\Delta:1:(1-\Delta)$	97,40	99,31	80,17	78,31	80,17		
		0,8	1,25	15.764	optimal	100	100	79,48	79,48	79,48	
					1:1:1	100	100	70,57	70,57	70,57	
					2:2:1	100	100	77,91	77,91	77,91	
					$\Delta:1:(1-\Delta)$	100	100	79,63	79,63	79,63	
	3		660	optimal	100	100	79,99	79,99	79,99		
				1:1:1	100	100	71,02	71,02	71,02		
				2:2:1	100	100	77,10	77,10	77,10		
				$\Delta:1:(1-\Delta)$	100	100	79,19	79,19	79,19		
	25%		0,5	1,25	5.212	optimal	97,43	99,54	79,64	77,52	79,63
						1:1:1	99,47	99,46	75,09	74,72	75,09
						2:2:1	98,98	99,00	78,37	77,66	78,37
						$\Delta:1:(1-\Delta)$	97,32	99,54	79,86	77,60	79,86
		3		258	optimal	97,12	99,09	79,79	77,57	79,79	
					1:1:1	98,50	98,80	76,64	75,54	76,64	
					2:2:1	95,74	95,53	74,59	72,39	74,59	
					$\Delta:1:(1-\Delta)$	94,96	98,45	78,90	75,13	78,90	
0,8		1,25		31.927	optimal	100	100	79,30	79,30	79,30	
					1:1:1	100	100	70,30	70,30	70,30	
					2:2:1	100	100	76,64	76,64	76,64	
					$\Delta:1:(1-\Delta)$	100	100	79,67	79,67	79,67	
		3	1.420	optimal	100	100	80,11	80,11	80,11		
				1:1:1	100	100	72,48	72,48	72,48		
				2:2:1	100	100	78,78	78,78	78,78		
				$\Delta:1:(1-\Delta)$	100	100	79,27	79,27	79,27		

Tabelle 6.1: Simulierte Güte in % der Tests basierend auf n_{semip}^* (ermittelt unter der Alternative $\beta_S = \beta_N$ bei einem nominalen Niveau von $\alpha = 2,5\%$ und für eine gewünschte Güte von 80% für den Nichtunterlegenheitstest). Die Fallzahlen n_{semip}^* wurden für $n_S:n_N:n_P$ nach dem optimalen Verhältnis und nach 1:1:1, 2:2:1 und $\Delta:1:(1-\Delta)$ aufgeteilt.

Wir stellen fest, dass der Nichtunterlegenheitstest sowohl für die variable Aufteilung aus [6.2.2](#) als auch für die Aufteilung $\Delta:1:(1-\Delta)$ die gewünschte Güte von 80% zufriedenstellend erreicht. Für die festen Aufteilungen 1:1:1 und 2:2:1 ist die simulierte Güte kleiner als die gewünschte; bei Parameterkonstellationen mit 1:1:1 und $\Delta = 0,8$ sogar um mehr als 10%. Die Güte des Tests für $H_0^{Ueb,S}$ ist für alle betrachteten Parameterkonstellationen meistens kleiner oder gleich der Güte des Tests für $H_0^{Ueb,N}$. Außerdem ist eine Abhängigkeit der Güte der Überlegenheitstests von Δ zu erkennen: je größer Δ , desto größer die Güte. Dies liegt daran, dass bei gleichen Parameterwerten aber größerem Δ mehr Fallzahlen nötig sind um bei dem Nichtunterlegenheitstest eine bestimmte Güte zu erreichen und dass die Überlegenheitshypothesen nicht von Δ abhängen.

Insgesamt ist die Güte der Überlegenheitstests wesentlich höher als die Güte des Nichtunterlegenheitstests und oft nahe 100%. Wegen dieser Tatsache stimmt die Güte des Nichtunterlegenheitstests gut mit der des Intersection-Union Tests überein.

Die simulierte Güte der Intersection-Union Tests für die festen Aufteilungen 1:1:1 und 2:2:1 ist 2% bis 11% geringer als die bei der variablen Aufteilung; für die Δ -abhängige Aufteilung hingegen ist die simulierte Güte nur bis zu 3% geringer und oft sogar fast identisch zu der bei der variablen Aufteilung.

Insgesamt halten wir also fest, dass es ausreicht, die Fallzahlplanung unter der Alternative $\beta_S = \beta_N \neq 0$ basierend auf dem Nichtunterlegenheitstest durchzuführen. Dabei spielt es keine Rolle, welcher der Überlegenheitstests verwendet wird.

Robustheit des Nichtunterlegenheitstests

Nun wollen wir anhand zweier Beispiele untersuchen, wie robust der Nichtunterlegenheitstest bezüglich des Erreichens der Güte und Einhalten des Niveaus ist, wenn die Überlebenszeiten in den Gruppen Weibull-verteilt mit unterschiedlichen Form-Parametern sind. Dabei wählen wir einmal, bezogen auf den Skalen-Parameter $\lambda_P = 1$ einen kleinen Skalen-Parameter von $\lambda_S = \lambda_N = 1,25$ und das andere mal einen großen von $\lambda_S = \lambda_N = 3$. Ansonsten verwenden wir in beiden Fällen dieselben Parameter, nämlich: Den gesamten Stichprobenumfang n_{semip}^* und dessen beste Aufteilung in die Gruppen ermitteln wir mit dem eben vorgestellten Algorithmus aus [Abschnitt 6.2.2](#). Als Toleranzwert haben wir jeweils $\epsilon = ((\beta_N - \Delta\beta_S)/(u_{1-\gamma} - u_{1-\alpha}))^2/100$ benutzt. Den Form-Parameter setzen wir auf $\vartheta = 1$, die Nichtunterlegenheitsgrenze auf $\Delta = 0,65$, das nominale Niveau auf $\alpha = 2,5\%$ und die gewünschte Güte auf $\gamma = 80\%$.

Basierend auf diesen Stichprobenumfängen erzeugen wir uns einen Datensatz mit Überlebenszeiten, die in jeder Gruppe folgenden Weibull-Verteilungen folgen: $T_S \sim Wb(\lambda_S, \vartheta_S)$,

$T_N \sim Wb(\lambda_N, \vartheta_N)$ und $T_P \sim Wb(\lambda_P, \vartheta_P)$. Nun simulieren wir die tatsächliche Fehlerwahrscheinlichkeit 1. Art und Güte unseres Nichtunterlegenheitstests für verschiedene Wahlen der Form-Parameter $\vartheta_S = \vartheta_N$ und ϑ_P . Insgesamt haben wir 10.000 Wiederholungen durchgeführt. Die Ergebnisse sind in Tabelle 6.2 dargestellt.

$\lambda_S = \lambda_N$	$\vartheta_S = \vartheta_N$	t.F.1	Güte
1,25	0,8	0,20	47,34
	0,9	0,72	64,74
	1	2,33	79,90
	1,1	7,02	90,04
	1,2	16,88	96,05
3	0,8	1,85	80,99
	0,9	2,43	79,43
	1	2,48	79,52
	1,1	2,79	78,86
	1,2	2,98	78,09

Tabelle 6.2: Simulierte tatsächliche Fehlerwahrscheinlichkeit 1. Art (t.F.1) in % zum nominalen Niveau von $\alpha = 2,5\%$ und simulierte Güte in % bei einer gewünschten Güte von 80% für den Nichtunterlegenheitstest bei $\Delta = 0,65$, $\lambda_P = 1$ und $\vartheta_P = 1$.

Wir betrachten zunächst die Ergebnisse für $\lambda_S = \lambda_N = 1,25$. Für $\vartheta_S = \vartheta_N = 0,8$; $0,9 < \vartheta_P = 1$ wird das Niveau nicht voll ausgeschöpft, dafür sinkt die Güte teilweise sogar drastisch ab. Hingegen wird für $\vartheta_S = \vartheta_N = 1,1$; $1,2 > \vartheta_P = 1$ das Niveau oft sogar erheblich überschritten. Für $\lambda_S = \lambda_N = 1,25$ ist dieses Verhalten wesentlich extremer als bei $\lambda_S = \lambda_N = 3$. Schon anhand dieser Beispiele sehen wir, dass je nach Wahl der Form-Parameter der Nichtunterlegenheitstest sehr robust, aber auch sehr empfindlich sein kann. Vor diesem Hintergrund ist auf jeden Fall ein Test auf proportionale Hazardfunktionen unverzichtbar, bevor mit der Auswertung eines Datensatzes begonnen wird (hierfür können beispielsweise die auf Seite 35 ff. vorgestellten Methoden verwendet werden). Um einen Einblick zu bekommen, wie stark die Annahme proportionaler Hazardfunktionen in der eben vorgestellten Simulationsstudie verletzt sein kann, haben wir in Abbildung 6.6 die theoretischen log-log-Überlebenszeitfunktionen abgebildet.

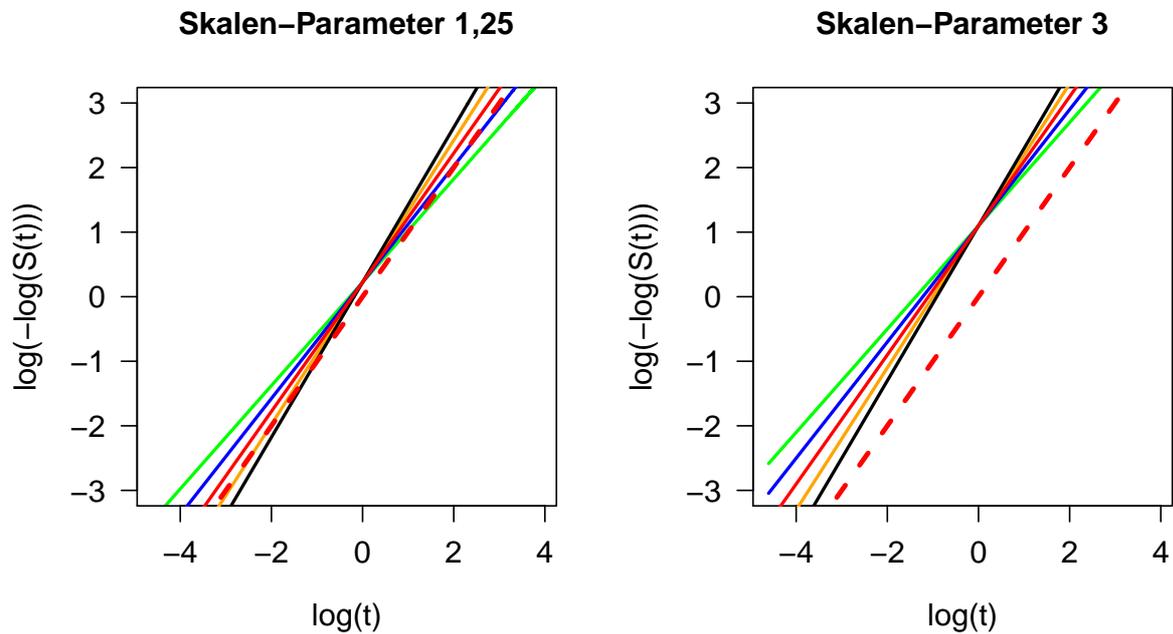


Abbildung 6.6: log-log-Überlebenszeitfunktionen der Weibull-Verteilung für die Form-Parameter **0,8**, **0,9**, **1**, **1,1** und **1,2** bei verschiedenen Skalen-Parametern (–). In jeder Abbildung ist außerdem die entsprechende Funktion für Skalen-Parameter 1 und Form-Parameter 1 (– – –) abgebildet.

Wenn die Annahme proportionaler Hazardfunktionen erfüllt ist, dann erwartet man, dass die log-log-Überlebenszeitkurve der Standard- und Novumgruppe (durchgezogene Linien) parallel zur log-log-Überlebenszeitkurve der Placebogruppe (gestrichelte Linie) verläuft. In der linken Graphik sehen wir deutlich, dass sich die gestrichelte mit jeder der durchgezogenen Linien schneidet. In der rechten Graphik schneiden die durchgezogenen Linien die gestrichelte nicht und verlaufen teilweise recht parallel zu ihr.

Damit haben wir in Kapitel **2** bis **6** eine semiparametrische Methode zur Planung und Auswertung von Nichtunterlegenheitsstudien entwickelt. Wie sich dieses Modell zu einem bestehenden parametrischen Modell unter der Annahme von exponentialverteilten Überlebenszeiten verhält, wird im nächsten Kapitel untersucht.

7 Vergleich mit einem parametrischen Verfahren

Wie in der Einleitung bereits erwähnt, gibt es in der Literatur Testverfahren für klinische Studien im Gold Standard Design, die auf einer bestimmten Verteilung der Überlebenszeiten basieren. Insbesondere wird in [Mielke et al. \[2008a\]](#) ein Nichtunterlegenheitstest für exponentialverteilte (entspricht einer Weibull-Verteilung mit Form-Parameter 1) Überlebenszeiten entwickelt. Hier wird also auch von proportionalen Hazardfunktionen ausgegangen. Außerdem wird dort eine Formel hergeleitet, mit der die optimale Anzahl der zu beobachtenden Personen und deren optimale Aufteilung in die Gruppen berechnet werden kann. Planen wir mit unserem Verfahren die Stichprobenumfänge für Weibull-verteilte Überlebenszeiten mit Form-Parameter 1, so gehen wir von derselben Verteilung der Überlebenszeiten wie [Mielke et al. \[2008a\]](#) aus.

Zunächst vergleichen wir die Fallzahlen beider Methoden unter der Annahme exponentialverteilter Überlebenszeiten und werden sehen, dass die ermittelten Zahlen kaum voneinander abweichen. Die relative asymptotische Effizienz des semiparametrischen Schätzers bezüglich des parametrischen ist nahe 1. Anschließend untersuchen wir, wie sich das parametrische Verfahren verhält, wenn die Daten Weibull-verteilt sind mit Form-Parameter $\vartheta \neq 1$. Dort werden wir feststellen, dass das parametrische Verfahren nicht robust bezüglich der Verletzung der Verteilungsannahme ist, während unsere Methode, obwohl wir die Fallzahlen von [Mielke et al. \[2008a\]](#) zugrunde gelegt haben, weiterhin das Niveau hält und die gewünschte Güte gut erzielt.

Schließlich greifen wir die Depressionsstudie auf und vergleichen auch hier beide Methoden miteinander. Wir erhalten verschiedene Ergebnisse, allerdings scheint der Stichprobenumfang zu gering für eine qualifizierte Aussage.

7.1 Parametrisches Verfahren

Wir stellen nun das parametrische Verfahren aus [Mielke et al. \[2008a\]](#) kurz vor und geben Formeln an, mit denen der gesamte Stichprobenumfang und dessen Aufteilung berechnet werden kann. Die Überlebenszeiten seien folgendermaßen verteilt:

$$T_S \sim \text{Exp}(\lambda_S) = \text{Wb}(\lambda_S, 1),$$

$$T_N \sim \text{Exp}(\lambda_N) = \text{Wb}(\lambda_N, 1),$$

$$T_P \sim \text{Exp}(\lambda_P) = \text{Wb}(\lambda_P, 1).$$

Die Hazardfunktionen haben in diesem Fall folgende Form:

$$h_k(t) = \lambda_k \quad \text{für } k = S, N, P.$$

Das Nichtunterlegenheits-Testproblem lautet:

$$H_0^{Nu,exp} : \log \lambda_N - \Delta \log \lambda_S + (1 - \Delta) \log \lambda_P \leq 0$$

$$H_1^{Nu,exp} : \log \lambda_N - \Delta \log \lambda_S + (1 - \Delta) \log \lambda_P > 0$$

mit der Teststatistik:

$$T^{Nu,exp} = \frac{\hat{\eta}}{\sqrt{\frac{1}{L_N} + \frac{\Delta^2}{L_S} + \frac{(\Delta-1)^2}{L_P}}}.$$

Dabei ist $\eta := \log \lambda_N - \Delta \log \lambda_S + (\Delta - 1) \log \lambda_P$ und $\hat{\eta}$ ist der Maximum-Likelihood-Schätzer von η , wobei $\hat{\lambda}_k := L_k/X^{(k)}$, L_k die Anzahl der Ereignisse und $X^{(k)}$ die Summe der Überlebens- und Zensierungszeiten aller Personen aus Gruppe k ist. Auf dem Rand der Hypothese (d.h. $\eta = 0$) ist $T^{Nu,exp}$ asymptotisch standardnormalverteilt. Die Hypothese $H_0^{Nu,exp}$ wird verworfen, falls $T^{Nu,exp} > u_{1-\alpha}$, wobei $u_{1-\alpha}$ das $(1 - \alpha)$ -Quantil der Standardnormalverteilung bezeichnet. Die optimale Personenanzahl beträgt:

$$n_{exp}^* := 2 \cdot \left(\frac{\Delta}{p_S} + \frac{1}{p_N} + \frac{1 - \Delta}{p_P} \right) \cdot \left(\frac{u_{1-\gamma} - u_{1-\alpha}}{\eta} \right)^2,$$

wobei $p_k := P(T_k < U_1, T_k < b)$.

Die optimale Aufteilung ist:

$$w_S^* = \frac{\Delta/p_S}{\Delta/p_S + 1/p_N + 1-\Delta/p_P}$$

$$w_N^* = \frac{1/p_N}{\Delta/p_S + 1/p_N + 1-\Delta/p_P}$$

$$w_P^* = 1 - w_S^* - w_N^*.$$

Es gilt also:

$$n_S^* : n_N^* : n_P^* = \Delta \frac{p_N}{p_S} : 1 : (1 - \Delta) \frac{p_N}{p_P}.$$

Um einen Eindruck über die Größe des benötigten Stichprobenumfanges zu erhalten, ist in Abbildung 7.1 für spezielle Parameterwahlen die insgesamt benötigte Anzahl an Personen dargestellt (ohne Berücksichtigung von zensierten Beobachtungen).

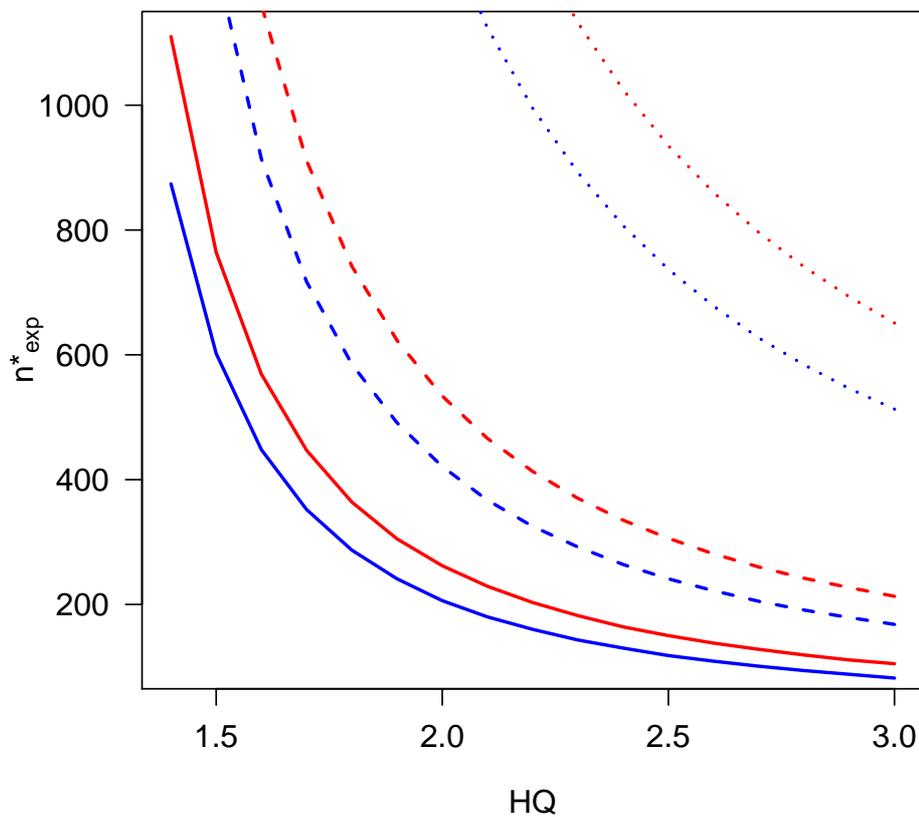


Abbildung 7.1: Insgesamt benötigter Stichprobenumfang n_{exp}^* für verschiedene Hazard Quotienten (HQ). Für $\Delta = 0,5$ (—), $\Delta = 0,65$ (---), $\Delta = 0,8$ (···), mit $\lambda_P = 1$, unter der Alternative $\lambda_S = \lambda_N$, bei einem nominalen Niveau von $\alpha = 2,5\%$ bzw. $\alpha = 5\%$ und einer gewünschten Güte von 80%.

7.2 Gegenüberstellung des parametrischen und semiparametrischen Verfahrens

7.2.1 Exponentialverteilte Überlebenszeiten

In diesem Abschnitt untersuchen wir, welche Stichprobenumfänge insgesamt bei dem parametrischen und dem semiparametrischen Verfahren notwendig sind, wenn eine Studie unter der Annahme von exponentialverteilten Überlebenszeiten geplant wird. Außerdem untersuchen wir, wie viel man an Güte verliert, wenn man den Stichprobenumfang mit dem parametrischen Verfahren plant, aber mit dem semiparametrischen auswertet. Schließlich gehen wir auf die asymptotische relative Effizienz (nach [van der Vaart \[1998\]](#), Kapitel 8) beider Schätzer ein.

Optimale Anzahl der insgesamt benötigten Patienten für beide Verfahren

Zunächst vergleichen wir die optimale Anzahl der insgesamt benötigten Patienten basierend auf dem parametrischen Verfahren (n_{exp}^*) mit denen basierend auf dem semiparametrischen (n_{semip}^*) bei Benutzung des Algorithmus aus Abschnitt 6.2.2. Wir haben mit beiden Methoden den erforderlichen Stichprobenumfang unter der Alternative $\beta_S = \beta_N$ für verschiedene Hazard Quotienten $HQ \in [1,25; 3]$, verschiedene Nichtunterlegenheitsgrenzen $\Delta \in \{0,5; 0,65; 0,8\}$, jeweils sowohl ohne Berücksichtigung von zensierten Beobachtungen (in diesem Fall ist $p_f^S, p_z^S = 10^{-20}\%$) als auch mit verschiedenen Anteilen an zufällig bzw. fest rechtszensierten Beobachtungen, bezüglich eines Niveaus von $\alpha = 2,5\%$ bzw. $\alpha = 5\%$ und einer Güte von $\gamma = 80\%$ berechnet. In Abbildung 7.2 sind die wesentlichen Ergebnisse bezüglich der relativen Abweichungen der beiden Stichprobenumfänge dargestellt.

Werden keine zensierten oder nur zufällig zensierte Beobachtungen bei der Fallzahlplanung berücksichtigt (vgl. Plots in der linken Spalte in Abbildung 7.2), so ist n_{semip}^* höher als n_{exp}^* . Allerdings nehmen die Abweichungen von n_{semip}^* bezüglich n_{exp}^* ab, je höher der Anteil der zufällig zensierten Beobachtungen gewählt wird. Fließt in die Berechnung der Stichprobenumfänge mit ein, dass fest zensierte Beobachtungen erwartet werden, so ist n_{semip}^* kleiner als n_{exp}^* (vgl. Plots in der rechten Spalte in Abbildung 7.2). Dieses unterschiedliche Verhalten der Stichprobenumfänge zueinander resultiert daraus, dass die optimale Aufteilung der Personen in den Gruppen bei beiden Methoden nicht identisch ist. Folglich sind die Wahrscheinlichkeiten für ein Ereignis im Verlauf der Studie (vgl. dazu Abschnitt 5.2) voneinander verschieden. Obwohl die Anzahl der erforderlichen Ereignisse bei dem parametrischen Verfahren gleich oder minimal geringer als die bei dem semiparametrischen ist, kann es also auftreten, dass n_{exp}^* kleiner oder größer als n_{semip}^* ist. Jedoch nimmt in allen be-

trachteten Fällen der Unterschied zwischen den beiden Stichprobenumfängen ab, je kleiner der Hazard Quotient und je größer Δ gewählt wird. Absolut gesehen unterscheiden sich die bei den betrachteten Parameterwahlen erforderlichen Stichprobenumfänge beider Verfahren allerdings nur unwesentlich (vgl. dazu Tabelle 7.1). In Tabelle 7.1 ist außerdem jeweils die simulierte Güte angegeben, die in allen betrachteten Fällen zufriedenstellend erreicht wird. Dies zeigt, dass die berechneten Stichprobenumfänge jeweils ausreichend sind.

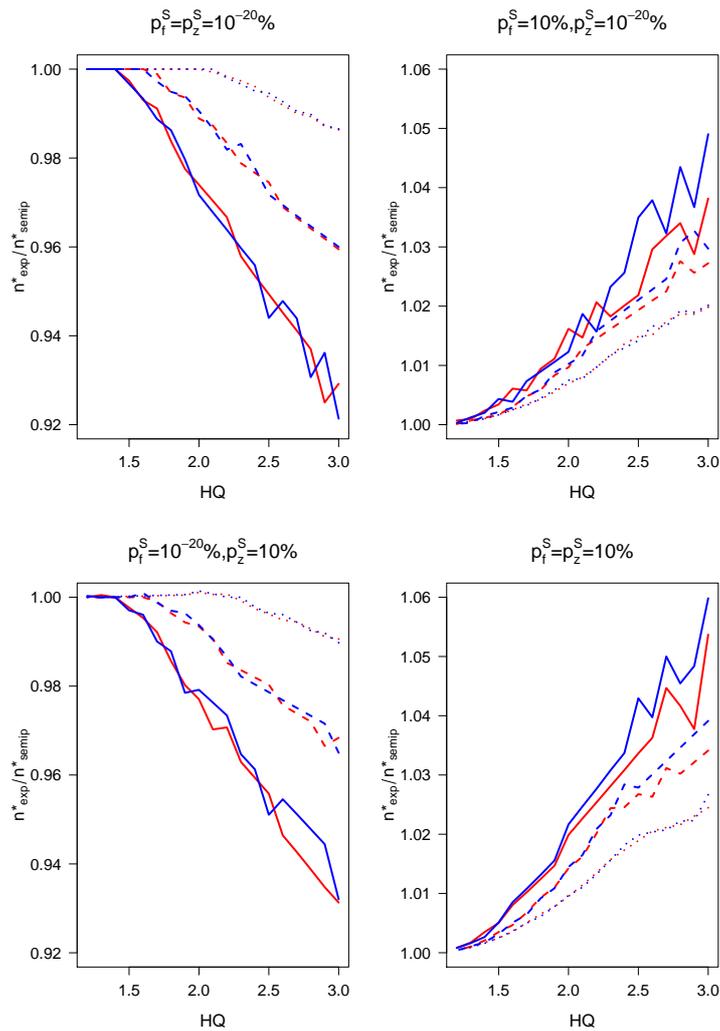


Abbildung 7.2: Abweichungen von n_{exp}^* und n_{semp}^* für verschiedene Nichtunterlegenheitsgrenzen $\Delta = 0,5$ (—), $\Delta = 0,65$ (---), $\Delta = 0,8$ (···) und einem nominalen Niveau von $\alpha = 2,5\%$ bzw. $\alpha = 5\%$ in Abhängigkeit des Hazard Quotienten (HQ).

$p_f = p_z$	α	Δ	HQ	n_{exp}^*	n_{semip}^*	$n_{semip}^* - n_{exp}^*$	
10 ⁻²⁰ %	2,5%	0,5	1,25	2.523 (79,35%)	2.523 (79,76%)	0	
			3	105 (78,94%)	113 (79,78%)	8	
	0,8	1,25	1,25	15.764 (80,13%)	15.764 (79,33%)	0	
			3	651 (79,87%)	660 (79,64%)	9	
	5%	0,5	1,25	1,25	1.987 (80,68%)	1.987 (79,60%)	0
				3	82 (77,92%)	89 (79,57%)	7
0,8		1,25	1,25	12.417 (79,80%)	12.417 (79,69%)	0	
			3	513 (79,76%)	520 (78,68%)	7	
10%	2,5%	0,5	1,25	3.223 (79,74%)	3.219 (80,25%)	-4	
			3	157 (80,58%)	149 (78,73%)	-8	
	0,8	1,25	1,25	19.880 (80,84%)	19.867 (80,23%)	-13	
			3	879 (80,01%)	858 (79,84%)	-21	
	5%	0,5	1,25	1,25	2.539 (79,71%)	2.536 (79,20%)	-3
				3	124 (79,74%)	117 (79,34%)	-7
0,8		1,25	1,25	15.659 (80,24%)	15.649 (80,39%)	-10	
			3	693 (80,40%)	675 (80,16%)	-18	

Tabelle 7.1: Absolute Abweichung von n_{exp}^* und n_{semip}^* . Die Werte in Klammern stellen die jeweils simulierte Güte dar (bei einer gewünschten Güte von 80%).

Die Fallzahlen, die mit dem parametrischen Verfahren ermittelt wurden, sind also gleichwertig zu denen, die mit dem semiparametrischen Verfahren berechnet wurden.

Vergleich von tatsächlichem Niveau und Güte beider Nichtunterlegenheitstests basierend auf dem Stichprobenumfang n_{exp}^*

Nun untersuchen wir, wie sich beide Tests in Hinblick auf die Einhaltung des Niveaus und das Erreichen der Güte verhalten. Die Stichprobenumfänge haben wir mit dem parametrischen Verfahren berechnet. Diese Fallzahlen und deren für das parametrische Verfahren optimale Aufteilung benutzen wir auch für den semiparametrischen Test. So können wir einen direkten Vergleich beider Nichtunterlegenheitstests durch Gegenüberstellung von simuliertem tatsächlichem Fehler 1. Art und Güte durchführen. Wir betrachten folgende Parameterwahlen: $HQ(S, P) \in \{1,25; 1,5; 3\}$, Nichtunterlegenheitsgrenzen $\Delta \in \{0,5; 0,65; 0,8\}$, verschiedene Anteile von zufälligen und fest rechtszensierten Beobachtungen $p_f^S, p_z^S \in \{10^{-20\%}; 10\%\}$ und Skalen-Parameter $\lambda_P = 1$. Das nominale Niveau beträgt 2,5% und die gewünschte Güte 80%. Für die Levelsimulationen haben wir λ_N gerade so bestimmt, dass $\beta_N = \Delta \beta_S$ gilt, wir uns also auf dem Rand der Hypothese befinden. Bei den Untersuchungen der Güte wurde $\beta_N = \beta_S$ gewählt. Wir haben jeweils 10.000 Wiederholungen durchgeführt. Die wesentlichen Ergebnisse sind in Tabelle 7.2 dargestellt.

p_z^S	ϑ	HQ	$p_f^S = 10^{-20\%}$				$p_f^S = 10\%$					
			n_{exp}^*	exp.		semip.		n_{exp}^*	exp.		semip.	
			t.F.1	Güte	t.F.1	Güte	t.F.1	Güte	t.F.1	Güte	t.F.1	Güte
10 ⁻²⁰ %	1	1,25	5.148	2,77	80,46	2,74	80,29	5.789	2,43	78,96	2,39	78,91
		1,5	1.559	2,52	79,55	2,45	79,13	1.777	2,39	79,57	2,41	79,46
		3	213	2,52	79,32	2,54	77,90	264	2,39	80,36	2,34	79,95
10%	1	1,25	5.745	2,48	80,24	2,49	80,02	6.535	2,64	79,39	2,64	79,28
		1,5	1.748	2,40	79,46	2,46	79,28	2.012	2,54	80,00	2,47	79,96
		3	245	2,42	79,72	2,47	78,44	303	2,68	80,79	2,59	80,48

Tabelle 7.2: Simulierter tatsächlicher Fehler 1. Art (t.F.1) und Güte in % für den exponential Nichtunterlegenheitstest und den semiparametrischen Nichtunterlegenheitstest zum nominalen Niveau von 2,5%, einer gewünschten Güte von 80% und $\Delta = 0,65$.

Wir sehen, dass beide Testverfahren das Niveau zufriedenstellend einhalten und die Güte erreichen. Allerdings ist die Güte bei dem parametrischen Verfahren geringfügig höher. Dies ist zu erwarten, da der Stichprobenumfang und dessen Aufteilung basierend auf dem parametrischen Verfahren optimal geplant wurde.

Asymptotische relative Effizienz

Aus Abschnitt 3.2 wissen wir, dass gilt

$$\sqrt{n} \left[\left(\widehat{\beta}_N - \Delta \cdot \widehat{\beta}_S \right) - \left(\beta_N - \Delta \cdot \beta_S \right) \right] \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N} \left(0, \sigma_{semip}^2 \right),$$

mit

$$\sigma_{semip}^2 = \frac{A_N - 2\Delta B + \Delta^2 A_S}{A_S A_N - B^2}$$

$$A_k = \lambda_k w_k \int_0^b e^{-t(u+2\lambda_k)} dt - \lambda_k^2 w_k^2 \int_0^b \frac{e^{-t(u+2\lambda_k)}}{w_S \lambda_S e^{-\lambda_S t} + w_N \lambda_N e^{-\lambda_N t} + w_P \lambda_P e^{-\lambda_P t}} dt$$

$$B = \lambda_S \lambda_N w_S w_N \int_0^b \frac{e^{-t(u+\lambda_S+\lambda_N)}}{w_S \lambda_S e^{-\lambda_S t} + w_N \lambda_N e^{-\lambda_N t} + w_P \lambda_P e^{-\lambda_P t}} dt.$$

und aus Theorem A.1 in [Mielke et al. \[2008b\]](#) ist bekannt, dass

$$\sqrt{n} \left[\left(\log \widehat{\lambda}_N - \Delta \log \widehat{\lambda}_S + (\Delta - 1) \log \widehat{\lambda}_P \right) - \left(\log \lambda_N - \Delta \log \lambda_S + (\Delta - 1) \log \lambda_P \right) \right] \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N} \left(0, \sigma_{exp}^2 \right),$$

mit

$$\sigma_{exp}^2 = \frac{(u - \lambda_S) \Delta^2}{w_S \lambda_S [e^{-b(\lambda_S+u)} - 1]} + \frac{u - \lambda_N}{w_N \lambda_N [e^{-b(\lambda_N+u)} - 1]} + \frac{(u - \lambda_P) (\Delta - 1)^2}{w_P \lambda_P [e^{-b(\lambda_P+u)} - 1]}$$

gilt. Strebt der Stichprobenumfang gegen unendlich, so konvergieren beide zentrierten Schätzer mit jeweils derselben Rate gegen eine asymptotische Normalverteilung mit jeweils Erwartungswert 0 und fester, aber unterschiedlicher Varianz. Als nächstes vergleichen wir deshalb die asymptotischen Varianzen beider Schätzer unter der Annahme von exponentialverteilten Daten. Dazu betrachten wir die asymptotische relative Effizienz ARE (nach [van der Vaart \[1998\]](#), Kapitel 8), d.h.

$$\text{ARE} := \frac{\sigma_{exp}^2}{\sigma_{semip}^2}.$$

Für viele verschiedene Parameterkonstellationen für $\lambda_S, \lambda_N, \lambda_P, u, \Delta, p_f^S, p_z^S$ haben wir ARE ausgerechnet. Der Übersichtlichkeit halber haben wir die einzelnen Tabellen nicht jeweils aufgeführt. Die Berechnungen haben ergeben, dass sich ARE dem Wert 1 annähert, je näher λ_S und λ_N beieinander liegen, oder je näher λ_S und λ_N an λ_P liegen, oder je größer Δ gewählt wird oder je mehr zensierte Beobachtungen erwartet werden (siehe beispielhaft erste Zeile von Tabelle 7.3). Selbst wenn die λ_k nicht eng beieinander liegen, zusätzlich Δ klein ist und wenig zensierte Beobachtungen vorliegen, ergeben sich in diesen extremen Situationen noch tolerierbare Werte (siehe beispielsweise zweite Zeile in Tabelle 7.3).

λ_S	λ_N	λ_P	Δ	$p_z^S = p_f^S$	ARE	1/ARE
1,1	1,1	1	0,8	25%	0,999995	1,000005
6	6	1	0,5	$10^{-20}\%$	0,787040	1,270583

Tabelle 7.3: ARE für $w_S:w_N:w_P = \Delta:1:(1 - \Delta)$ und bei $w_N = 0,5$.

Zusammenfassend lässt sich feststellen, dass beide Verfahren gleichwertig sind, wenn die Daten Weibull-verteilt mit Form-Parameter $\vartheta = 1$ sind.

7.2.2 Weibull-verteilte Überlebenszeiten (Form-Parameter $\vartheta \neq 1$)

Nun untersuchen wir, wie sich beide Testverfahren verhalten, wenn Weibull-verteilte Daten mit Form-Parameter $\vartheta \neq 1$ vorliegen, die Planung der Studie aber mit dem parametrischen Verfahren erfolgte. Dazu simulieren wir Niveau und Güte und stellen sie einander gegenüber. Um zu untersuchen wie robust der parametrische Nichtunterlegenheitstest in diesem Fall ist, lassen wir das Ergebnis eines vorgeschalteten Tests, der die Daten auf Exponentialverteilung prüft, außer Acht und betrachten auch stark von 1 abweichende Form-Parameter. Wir berechnen n_{exp}^* basierend auf folgenden Parameterwahlen: $HQ(S, P) \in \{1,25; 1,5; 3\}$, Nichtunterlegenheitsgrenzen $\Delta \in \{0,5; 0,65; 0,8\}$, verschiedene Anteile von zufälligen und fest rechtszensierten Beobachtungen $p_f^S, p_z^S \in \{10^{-20}\%; 10\%\}$ und Skalen-Parameter $\lambda_P = 1$. Das nominale Niveau beträgt 2,5% und die gewünschte Güte 80%. Für die Levelsimulationen haben wir λ_N gerade so bestimmt, dass $\beta_N = \Delta \beta_S$ gilt, wir uns also auf dem Rand der Hypothese befinden. Bei den Untersuchungen der Güte wurde $\beta_N = \beta_S$ gewählt. Wir haben jeweils 10.000 Wiederholungen durchgeführt. Die wesentlichen Ergebnisse sind in Tabelle 7.4 und in Abbildung 7.3 dargestellt.

p_z^S	ϑ	HQ	$p_f^S = 10^{-20}\%$				$p_f^S = 10\%$					
			n_{exp}^*	exp.		semip.		n_{exp}^*	exp.		semip.	
			t.F.1	Güte	t.F.1	Güte	t.F.1	Güte	t.F.1	Güte	t.F.1	Güte
10 ⁻²⁰ %	0,5	1,25	5.148	17,21	94,90	2,59	80,10	5.789	6,00	86,73	2,35	75,98
		1,5	1.559	16,89	94,52	2,45	79,41	1.777	6,01	87,56	2,47	77,38
		3	213	14,65	93,60	2,60	78,40	264	5,14	87,94	2,54	81,08
	1,5	1,25	5.148	0,24	44,11	2,26	79,83	5.789	0,66	61,45	2,70	82,81
		1,5	1.559	0,21	44,80	2,43	78,82	1.777	0,52	63,39	2,38	81,30
		3	213	0,22	42,94	2,70	78,51	264	1,65	69,96	2,54	77,42
10%	0,5	1,25	5.745	16,29	93,86	2,35	79,44	6.535	6,32	87,31	2,45	77,37
		1,5	1.748	16,13	93,85	2,38	79,85	2.012	5,86	87,50	2,36	77,99
		3	245	13,22	92,35	2,63	78,56	303	5,28	87,67	2,62	81,68
	1,5	1,25	5.745	0,39	50,16	2,48	79,67	6.535	0,84	65,86	2,49	82,18
		1,5	1.748	0,28	50,67	2,29	78,83	2.012	0,91	67,43	2,33	80,92
		3	245	0,50	54,15	2,41	79,18	303	1,81	70,89	2,54	77,05

Tabelle 7.4: Simulierte tatsächliche Fehlerwahrscheinlichkeit 1. Art (t.F.1) und Güte in % für den parametrischen und den semiparametrischen Nichtunterlegenheitstest; zum nominalen Niveau von 2,5%, einer gewünschten Güte von 80% und $\Delta = 0,65$.

Für $\vartheta < 1$ wird bei dem parametrischen Verfahren das Niveau erheblich überschritten, aber eine höhere Güte erreicht. Genau umgekehrt wird für $\vartheta > 1$ das Niveau nicht voll ausgeschöpft, dafür sinkt die Güte drastisch ab. In Abbildung 7.3 ist dies mit roten Strichen auf den blauen Kurven markiert. Das semiparametrische Verfahren hingegen erreicht selbst basierend auf den nach Mielke et al. [2008a] ermittelten Fallzahlen bei allen hier betrachteten Verteilungen und für alle Parameterwahlen das Niveau und die gewünschte Güte zufriedenstellend, auch für sehr große Form-Parameter (vgl. Abbildung 7.4).

Ermittelt man hingegen die Fallzahlen innerhalb unseres Modells mit dem in Abschnitt 6.2.2 vorgestellten Algorithmus, erhalten wir hier bei einem Hazard Quotienten von 1,25 geradezu identische Fallzahlen in den einzelnen Gruppen und auch insgesamt. Bei einem Hazard Quotienten von 3 benötigen wir hingegen für unser Verfahren knapp 5% höhere Fallzahlen insgesamt als bei dem parametrischen (vgl. dazu auch Graphik links oben in Abbildung 7.2); auch gibt es eine leichte Verschiebung innerhalb der Gruppen. Mit diesen Fallzahlen wird dann eine Güte von 80% erreicht.

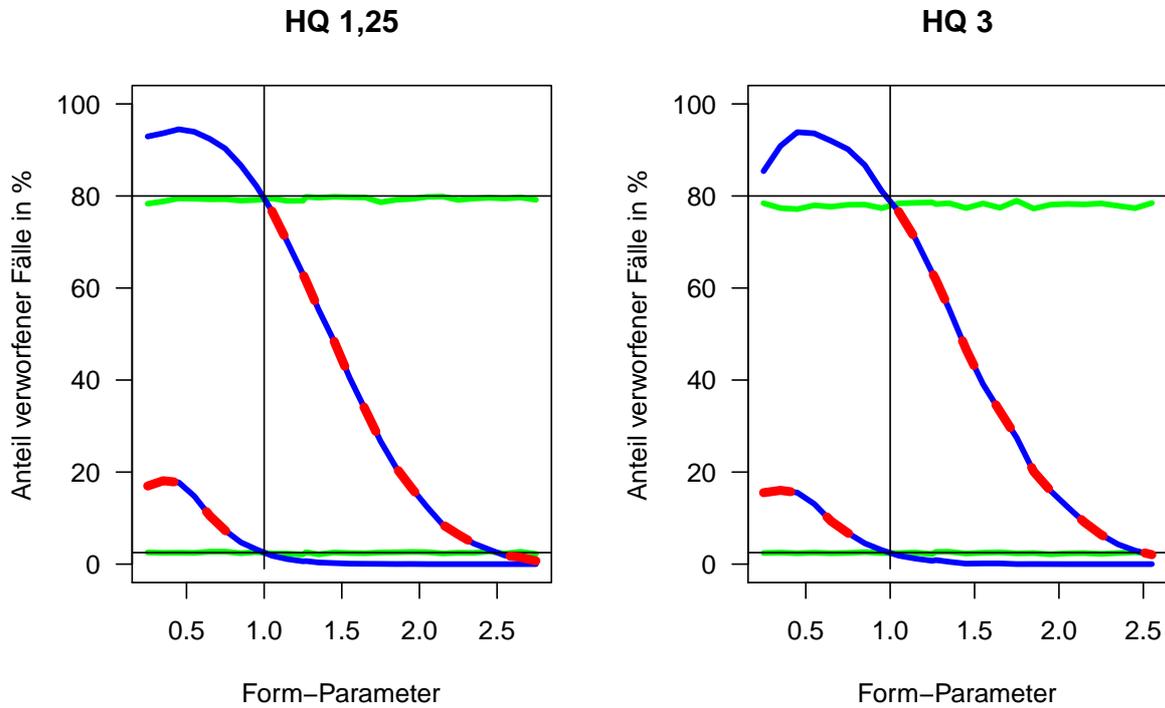


Abbildung 7.3: Simulierte tatsächliche Fehlerwahrscheinlichkeit 1. Art und simulierte Güte des **parametrischen** (**starke Abweichungen** zu nominalen Werten) und des **semiparametrischen** Nichtunterlegenheitstests zum nominalen Niveau von 2,5%, einer gewünschten Güte von 80% und $\Delta = 0,65$.

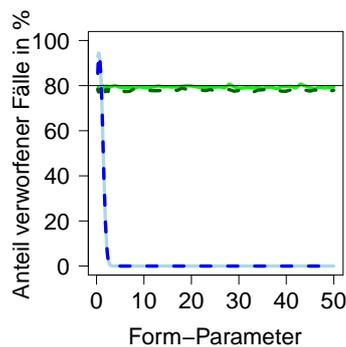


Abbildung 7.4: Simulierte Güte des **semiparametrischen** und den **parametrischen** Nichtunterlegenheitstests für $HQ=1,5$ (—) und $HQ=3$ (---).

Um einen Einblick zu bekommen, wie sich die Form der Dichtefunktionen der Weibull-Verteilung bei unterschiedlichen Form-Parametern verändert, haben wir in [Abbildung 7.5](#) verschiedene Dichtefunktionen abgebildet.

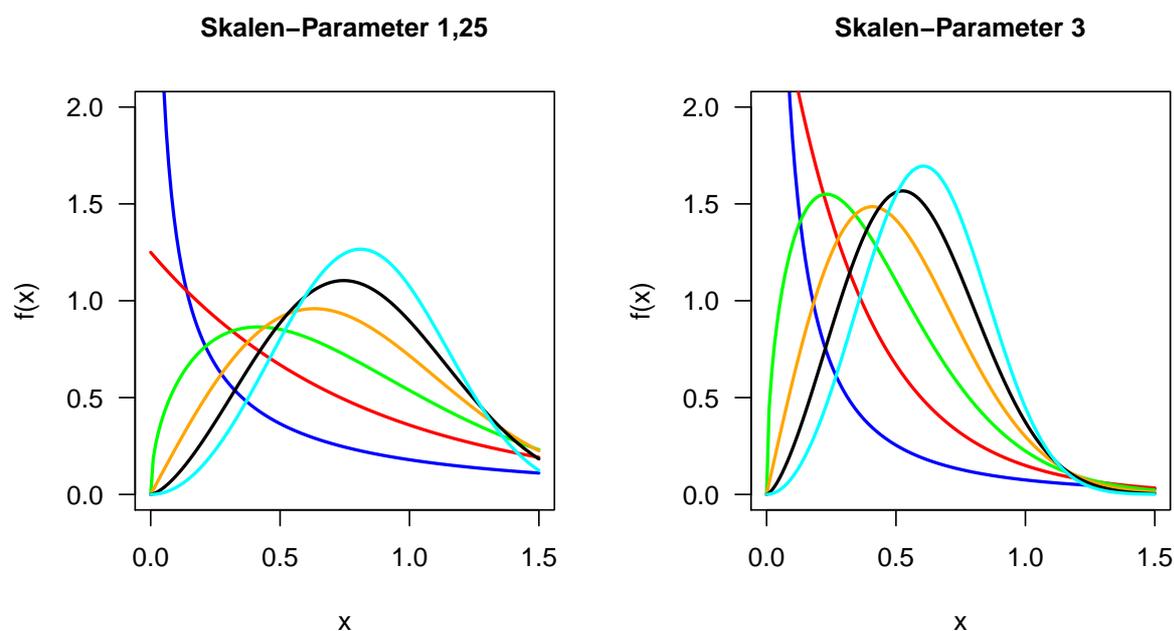


Abbildung 7.5: Dichtefunktionen der Weibull-Verteilung für verschiedene Form-Parameter $\vartheta \in \{0,5, 1, 1,5, 2, 2,5, 3\}$.

Zusammenfassend lässt sich feststellen, dass in diesem Vergleich das parametrische Verfahren versagt, wenn die Verteilung der Daten bereits geringfügig von einer Exponentialverteilung abweicht. Unser semiparametrisches Verfahren hingegen ist hier unempfindlich und liefert weiterhin zuverlässige Aussagen.

7.2.3 Auswertung der Depressionsstudie

Die Depressionsstudie wurde ebenfalls von [Mielke et al. \[2008a\]](#) untersucht, so dass wir anhand dieses Datensatzes beide Verfahren noch einmal konkret vergleichen können.

Modell 3: Hier werden Weibull-Verteilungen mit Form-Parameter $\vartheta = 1$ an die Daten angepasst. Die aus den Daten mit der Maximum-Likelihood-Methode geschätzten Parameter lauten:

$$\hat{\lambda}_{S,3} = 0,01193, \quad \hat{\lambda}_{N,3} = 0,01476 \quad \text{und} \quad \hat{\lambda}_{P,3} = 0,01113$$

und das Testen der verschiedenen Hypothesen führt zu den p-Werten in Tabelle 7.5 bzw. Tabelle 7.6. Dabei haben wir die p-Werte unseres semiparametrischen Ansatzes als rechte Spalte angefügt.

Hypothesen	exp.	semip.
$H_0^{Ueb,S}$	33,34%	37,09%
$H_0^{Ueb,N}$	3,88%	5,17%

Tabelle 7.5: P-Werte der Überlegenheitstests.

Wie den p-Werten in der zweiten Zeile von Tabelle 7.5 zu entnehmen ist, kann die Hypothese $H_{0,N}^{Ueb}$ zum 5% Niveau bei dem parametrischen Verfahren abgelehnt werden, während dies bei dem semiparametrischen Verfahren nicht der Fall ist. Die beiden Auswertungsmethoden führen also zu unterschiedlichen Ergebnissen.

Für das parametrische Verfahren führen wir nun den Nichtunterlegenheitstest durch. Aus Vergleichszwecken tabellieren wir auch hier die entsprechenden p-Werte für das semiparametrische Verfahren, obwohl diese Ergebnisse irrelevant für die Zulassung des neuen Medikamentes sind, da weder die Überlegenheit von Novum noch von Standard gegenüber Placebo zum 5% Niveau angenommen werden kann. Die Ergebnisse sind in Tabelle 7.6 zusammengefasst.

Hypothesen	exp.	semip.
$H_0^{Nu}, \Delta = 0,5$	1,83%	2,39%
$H_0^{Nu}, \Delta = 0,65$	1,91%	2,44%
$H_0^{Nu}, \Delta = 0,8$	2,51%	3,04%

Tabelle 7.6: P-Werte der Nichtunterlegenheitstests für verschiedene Nichtunterlegenheitsgrenzen Δ .

Hier sehen wir, dass bei beiden Verfahren zum 5% Niveau alle Nichtunterlegenheitshypothesen für die betrachteten Nichtunterlegenheitsgrenzen ablehnt werden können. Basierend auf dem parametrischen Verfahren dürfte das neue Medikament also auf dem Markt zugelassen werden, wenn als Überlegenheitshypothese $H_0^{Ueb,N}$ ausgewählt worden wäre.

Beide Verfahren gehen von proportionalen Hazardfunktionen aus; das Verfahren von [Mielke et al. \[2008a\]](#) basiert zusätzlich auf der Exponentialverteilung.

Um zu entscheiden, ob einem Modell der Vorzug gegeben werden kann, kommen wir auf unser Modell 2 von Seite 19 ff. zurück. Dort hatten wir Weibull-Verteilungen mit demselben

Form-Parameter an die Daten angepasst und erhielten folgende Ergebnisse:

$$\hat{\lambda}_{S,2} = 0,00418, \quad \hat{\lambda}_{N,2} = 0,00525, \quad \hat{\lambda}_{P,2} = 0,00391 \quad \text{und} \quad \hat{\vartheta}_2 = 1,2743.$$

Anhand von P-Plots soll in Abbildung 7.6 die Anpassung der empirischen Verteilungen an die mit obigen $\hat{\lambda}_{S,3}$, $\hat{\lambda}_{N,3}$ und $\hat{\lambda}_{P,3}$ ermittelten Exponentialverteilungen bzw. den Ergebnissen aus Modell 2 beurteilt werden.

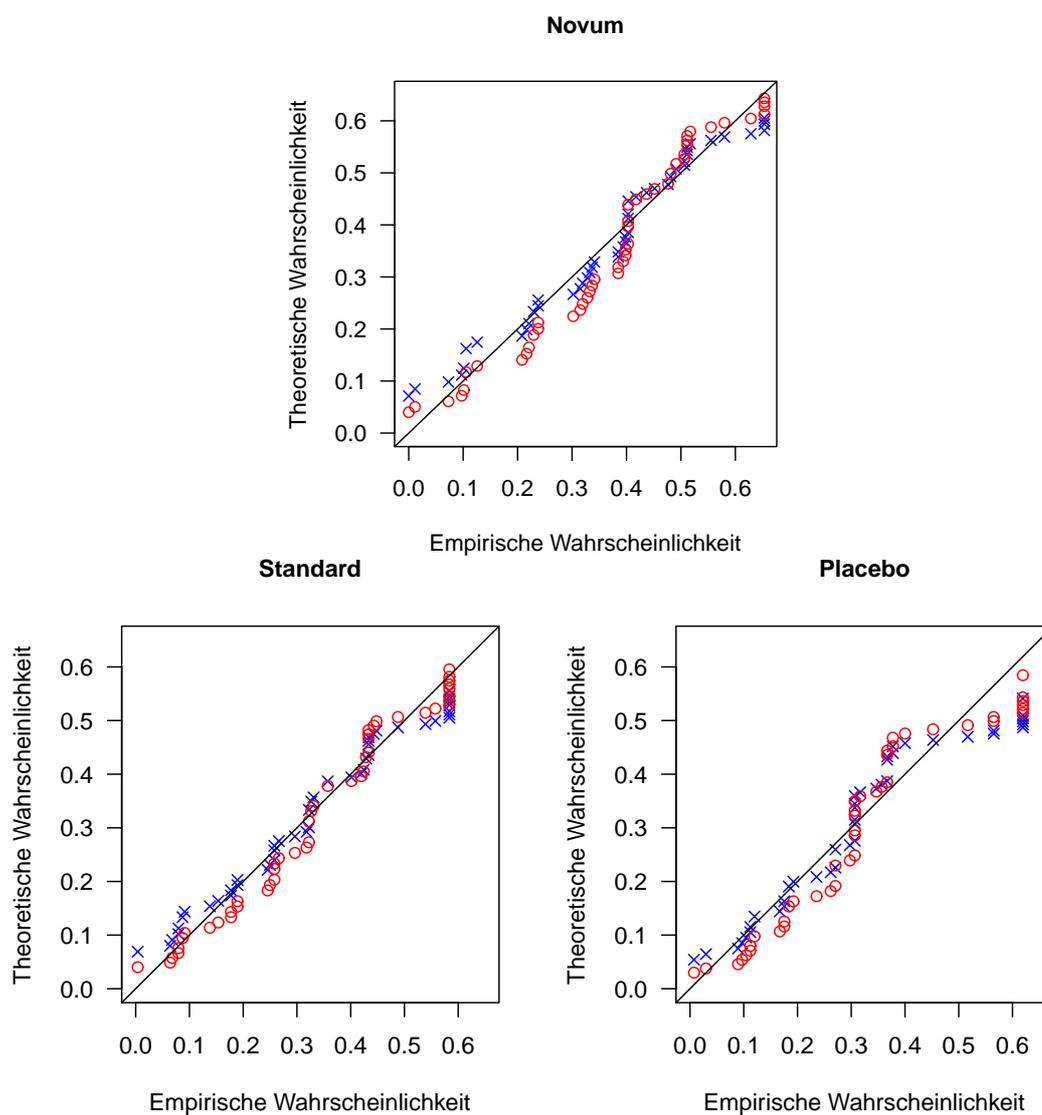


Abbildung 7.6: P-Plots für **Modell 2** rote Kreise, **Modell 3** blaue Kreuze

Die P-Plots lassen keine Schlüsse zu, ob die Kreuze oder die Kreise näher an der Ursprungsgeraden liegen. Deshalb vergleichen wir die beiden Modelle mittels des Likelihood-Quotienten-Tests. Sei $\Theta := \{\theta = (\lambda_S, \lambda_N, \lambda_P, \vartheta) : \lambda_S, \lambda_N, \lambda_P, \vartheta \in \mathbb{R}_{>0}\}$ der Parameterraum und $\Theta_0 := \{\theta = (\lambda_S, \lambda_N, \lambda_P, \vartheta) : \lambda_S, \lambda_N, \lambda_P \in \mathbb{R}_{>0}, \vartheta = 1\}$ der Hypothesenraum. Mit \widehat{L}_j sei der maximale Wert der log-Likelihoodfunktion von Modell j bezeichnet, $j = 2, 3$. Die beiden Modelle können nun anhand ihrer Werte von $-2 \log \widehat{L}_j$ miteinander verglichen werden. Als Niveau wählen wir 5%. Dazu betrachten wir die Teststatistik $-2 \log \widehat{L} = 2 \log \widehat{L}_2 - 2 \log \widehat{L}_3$. Unter $H_0 : \theta \in \Theta_0$ ist $-2 \log \widehat{L} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \chi_{\dim \Theta - \dim \Theta_0}^2 = \chi_1^2$.

Für den Datensatz ist $-2(\log \widehat{L}_2 - \log \widehat{L}_3) = 21,76$. Da der p-Wert $3,1 \cdot 10^{-4}\%$ beträgt, kann die Hypothese zum 5% Niveau verworfen werden. Das Modell 2 beschreibt die Daten also signifikant besser als das Modell 3. Von daher wäre unserem semiparametrischen Modell gegenüber dem parametrischen der Vorzug zu geben.

Abschließend betrachten wir Simulationen der Güte für die gleichen Stichprobenumfänge und deren Aufteilungen der Personen in die Gruppen, gleiche Zensierungsanteile und die gleiche Studiendauer wie in der Depressionsstudie. Einmal führen wir die Simulationen in Modell 2 durch unter Auswertung mit dem semiparametrischen Verfahren und auch in Modell 3 unter Auswertung mit dem parametrischen Ansatz; jeweils sowohl zum 2,5% als auch zum 5% Niveau. Die Ergebnisse sind in Tabelle 7.7 dargestellt.

Hypothesen	2,5% Niveau		5% Niveau	
	exp.	semip.	exp.	semip.
$H_0^{Nu}, \Delta = 0,5$	50,74%	55,97%	63,91%	67,99%
$H_0^{Nu}, \Delta = 0,65$	49,44%	56,16%	62,47%	68,10%
$H_0^{Nu}, \Delta = 0,8$	47,27%	52,94%	58,12%	64,04%

Tabelle 7.7: Simulierte Güte des parametrischen und semiparametrischen Nichtunterlegenheitstests für verschiedene Nichtunterlegenheitsgrenzen Δ bezüglich des 2,5% bzw. 5% Niveaus.

Hier ist auffällig, dass die Güte deutlich kleiner ist als die sonst von uns gewünschten 80%. Somit sind die Fallzahlen zu gering um eine falsche Nullhypothese mit der üblicherweise gewünschten Wahrscheinlichkeit von 80% zu verwerfen. Jedoch ist die Güte bei dem semiparametrischen Modell stets höher als bei dem parametrischen. Also ist ein Verwerfen einer falschen Hypothese bei dem semiparametrischen Verfahren wahrscheinlicher.

8 Erweiterungen des Modells

In den vorangegangenen Kapiteln haben wir ein neues statistisches Verfahren vorgestellt. Oft ergeben sich in der Praxis konkrete Frage-/Problemstellungen, die eine Anpassung unseres Modells erfordern. In den folgenden beiden Abschnitten möchten wir Anregungen geben, wie unser Verfahren für zwei Fragestellungen modifiziert werden kann. Zum einen betrachten wir variable Nachbeobachtungszeiten der Patienten und zum anderen wollen wir zusätzliche Kovariablen berücksichtigen.

8.1 Variable Nachbeobachtungszeit

In unserem Modell war die maximale Verweildauer für alle Patienten gleich lang gewählt. Dies hat zur Folge, dass es kein festes Studienenddatum für alle Patienten gibt, da im Allgemeinen die Patienten nicht alle zum gleichen Zeitpunkt in die Studie eintreten. Oft besteht das Interesse, dass die Studie zu einem festen Zeitpunkt beendet werden soll. Wenn weiterhin nicht alle Patienten zu einem festen Zeitpunkt eintreten, hat dies zur Folge, dass die maximalen Verweildauern vom Eintrittszeitpunkt in die Studie abhängen. In diesem Fall modifizieren wir unser Modell so, dass wir weiterhin von einem festen gemeinsamen Studienbeginn ausgehen, jetzt aber ein anderer fester Zensierungsmechanismus zu verwenden ist. D.h. die unterschiedlichen Eintritte „vorne“ berücksichtigen wir durch unterschiedliche feste Zensierungen „hinten“.

Wir nehmen ohne Einschränkung an, dass die Aufnahmezeit die Länge $a \cdot b$ hat, mit $a \in (0, 1]$ und b die Zeit vom Beginn der Aufnahmezeit bis zum Studienendzeitpunkt. Je kleiner a gewählt wird, desto kürzer ist die Aufnahmezeit, im Fall $a = 1$ werden über den gesamten Zeitraum Patienten rekrutiert. Somit beträgt die maximale Verweildauer ab dem

Eintrittszeitpunkt $g \in [0, a \cdot b]$ noch $b - g$. Wir nehmen an, dass die Patienten während der Aufnahmezeit gleichverteilt in die Studie eintreten und zwar sowohl bezogen auf die Eintrittszeitpunkte als auch bezüglich der Zuteilung zu den einzelnen Gruppen unter Berücksichtigung der Verhältnisse. Letzteres kann durch Blockrandomisierung wie z.B. in [Schulz and Grimes \[2002\]](#) beschrieben, erreicht werden.

In Kapitel 2-6 bestanden die beobachteten Daten für jede Person $i \in \{1, \dots, n\}$ aus dem Paar $(\min\{T_i, U_i, b\}, \delta_i)$ (vgl. Seite 9). Jetzt bestehen sie für jede Person i aus dem Paar $(\min\{T_i, U_i, G_i\}, \tilde{\delta}_i)$. Dabei sei $G_i := b - A_i$, wobei $A_i \sim \mathcal{U}[0, a \cdot b]$, also gleichverteilt ist. Die Verteilungsfunktion von G_i lautet

$$F_{G_i}(t) := 1 - F_{A_i}(b - t) = \begin{cases} 0 & : t < b - a \cdot b \\ 1 - \frac{b-t}{a \cdot b} & : t \in [b - a \cdot b, b] \\ 1 & : t > b \end{cases}$$

und der Zensierungsindikator

$$\tilde{\delta}_i := \mathbb{1}_{\{T_i < \min\{U_i, G_i\}\}} = \begin{cases} 1 & : \text{bei Person } i \text{ tritt während der Studie das Zielereignis ein} \\ 0 & : \text{Person } i \text{ ist zensiert.} \end{cases}$$

In Tabelle 8.1 haben wir in dem Modell aus den Kapiteln 2-6 mit den in Kapitel 6 vorgestellten Algorithmen die Fallzahlen insgesamt berechnet und die simulierte Güte des Nichtunterlegenheitstests aufgelistet.

Aufteilung	n_S	n_N	n_P	n	Güte in %
optimal	213	423	233	869	79,69
1:1:1	323	323	323	969	80,06
1:2:1	218	436	218	872	80,01

Tabelle 8.1: Optimale Fallzahlen bei $HQ = 1,5$, $\lambda_P = 1$, $\vartheta = 0,5$, $\Delta = 0,5$, einem Anteil von fest und zufällig zensierten Beobachtungen in der Standardgruppe von 5%, $b = 3,6$ unter der Alternative $\beta_S = \beta_N$, bezüglich eines nominalen Niveaus von 2,5%, einer gewünschten Güte von 80% und simulierte Güte des Nichtunterlegenheitstests.

In einem Studiendesign mit variabler maximaler Verweildauer erwarten wir mehr zensierte Beobachtungen. Deshalb sind hier entweder die Stichprobenumfänge aus Tabelle 8.1 zu erhöhen, wenn bei gleichen Parameterwahlen die gewünschte Güte weiterhin 80% betragen

soll, oder die Güte verringert sich bei Verwendung der Stichprobenumfänge aus Tabelle 8.1. Bezüglich der optimalen Aufteilung der Patienten in die Behandlungsgruppen erwarten wir im Allgemeinen dieselben Verhältnisse wie in dem Design mit gleich langer maximaler Verweildauer für alle Patienten, wenn die Patienten gleichverteilt in die Studie eintreten. Basierend auf den Fallzahlen aus Tabelle 8.1 simulieren wir dazu für verschieden lange Aufnahmephasen die Güte des Nichtunterlegenheitstests. Die Ergebnisse sind in Tabelle 8.2 aufgelistet.

Aufteilung	a=0,2	a=0,5	a=0,8
optimal	79,21	78,19	76,76
1:1:1	79,04	77,83	76,25
1:2:1	79,17	78,71	76,36

Tabelle 8.2: Simulierte Güte des Nichtunterlegenheitstests in % bei gleichverteiltem Studieneintritt in der Aufnahmephase $[0; a \cdot 3,6]$.

Wie zu erwarten, sinkt die Güte für länger gewählte Aufnahmephasen. Da sich allerdings die Güte für verschieden lange Aufnahmephasen „gleichmäßig“ bei allen Aufteilungen verringert, können wir davon ausgehen, dass die optimale Aufteilung in dem neuen Setting beibehalten werden kann.

8.2 Kovariablen

Im Allgemeinen gibt es neben den verschiedenen Medikamenten als primäre Variable noch andere Größen, die die Zielvariable einer klinischen Studie beeinflussen. In der ICH-Guideline E9 ICH [1998] und den „Points to consider on adjustment for baseline covariates“, CPMP [2003] wird deshalb empfohlen, die wichtigsten Kovariablen, die einen Einfluss auf die Zielvariable haben, bereits im Vorfeld zu identifizieren und im Studienprotokoll zu erfassen. Insbesondere ist hierbei der Einfluss des Baseline-Wertes der Zielvariable (in der vorliegenden Depressionsstudie der HAMD₁₇-Wert zu Beginn der Studie) oder auch bei Studien, die in verschiedenen Kliniken durchgeführt werden, der Einfluss der verschiedenen Kliniken zu beachten. Bereits bei der Randomisierung können die einzelnen Ausprägungen der Kovariablen gleichmäßig auf die Medikamentengruppen aufgeteilt werden (stratifizierte Randomisierung, vgl. etwa Schulz and Grimes [2002]). Im Allgemeinen wird für die Fallzahlplanung nur die primäre Variable berücksichtigt (vgl. ICH-Guideline E9 ICH [1998]). Bei der Auswertung der Studie wird der Behandlungseffekt üblicherweise um Kovariablen

adjustiert. Das Cox-Modell kann auch eingesetzt werden, wenn mehrere Kovariablen berücksichtigt werden sollen. Die Hazardfunktion für Person $i \in \{1, \dots, n\}$ im Cox-Modell sind dann beispielsweise gegeben durch

$$h_i(t) = h_0(t)e^{\beta Z_i + \alpha_K K_i + \alpha_H H_i},$$

bei j verschiedenen Kliniken sei $K_i := (K_{1,i}, \dots, K_{j-1,i})^T$ mit

$$K_{l,i} := \begin{cases} 1 & \text{Person } i \text{ ist in Klinik } l \\ 0 & \text{sonst,} \end{cases}$$

für $l \in \{1, \dots, j-1\}$ und $\alpha_K \in \mathbb{R}^{j-1}$ der Vektor der Regressionskoeffizienten. Mit H_i sei der HAMD₁₇-Wert zu Beginn der Studie und mit $\alpha_H \in \mathbb{R}$ der Regressionskoeffizient bezeichnet.

Alternativ kann das stratifizierte Cox-Modell benutzt werden. Hier werden die Personen entsprechend ihrer Zugehörigkeit zu den verschiedenen Ausprägungen der Kovariablen in disjunkte Gruppen aufgeteilt. Die disjunkten Gruppen stellen dann die Strata dar. In diesem Modell sind die Regressionskoeffizienten β in jedem Stratum dieselben, jedoch können die Baseline-Hazardfunktionen jeweils verschieden sein. Bei M Strata lauten die Hazardfunktionen für Person $i \in \{1, \dots, n\}$ dann:

$$h_{m,i}(t) = h_{0,m}(t)e^{\beta Z_i}$$

für $m \in \{1, \dots, M\}$ und β, Z_i wie auf Seite 11 bzw. in (2.2) definiert. Der Regressionskoeffizient β wird geschätzt, indem eine allgemeinere Partial-Likelihoodfunktion als die in (3.1) verwendet wird:

$$L(\beta) = \prod_{m=1}^M L_m(\beta).$$

Dabei ist jeder Faktor $L_m(\beta) = \prod_{l=1}^L \frac{e^{\beta Z(l)}}{\sum_{r \in R_t l} e^{\beta Z_r}}$ eine Partial-Likelihoodfunktion wie in (3.1), wobei jeweils nur Daten von Personen $l \in \{1, \dots, n\}$ aus Stratum $m \in \{1, \dots, M\}$ eingehen.

9 Zusammenfassung und Diskussion

3-armige Nichtunterlegenheitsstudien haben in den letzten Jahren zunehmend an Bedeutung gewonnen. Vor diesem Hintergrund wurde der Entwurf einer „Guidance for Industry on Non-Inferiority Clinical Trials“ der FDA [2010] oder auch die „Guideline on the choice of the non-inferiority margin“, CHMP [2005] entwickelt (vgl. [Brown et al. \[2006\]](#)).

Parametrische Ansätze für die Planung und Auswertung von 3-armigen Nichtunterlegenheitsstudien sind bereits in der Literatur zu finden. So können für Studien unter der Annahme von exponential- ([Mielke et al. \[2008a\]](#)), normal- ([Pigeot et al. \[2003\]](#), [Hasler et al. \[2008\]](#)), binomial- ([Kieser and Friede \[2007\]](#), [Mielke and Munk](#)) oder Poisson-verteilten ([Mielke and Munk](#)) Überlebenszeiten die erforderlichen Stichprobenumfänge und deren optimale Aufteilung in die Behandlungsgruppen geplant werden.

Vor diesem Hintergrund wird in dieser Arbeit ein allgemeiner semiparametrischer Ansatz entwickelt. Dabei gehen wir davon aus, dass die Hazardfunktionen für die 3 Behandlungsgruppen proportional zueinander sind. Dies stellt die zentrale Annahme in unserem Ansatz dar und ermöglicht die Anwendung des Cox-Regressions-Modells. Die interessierenden Regressionskoeffizienten der Hazardfunktionen werden dabei mit der Partial-Maximum-Likelihood-Methode geschätzt. Basierend auf dieser Methode entwickeln wir ein Testverfahren und stellen einen Algorithmus vor, mit dem der insgesamt erforderliche Stichprobenumfang, dessen optimale Aufteilung in die Behandlungsgruppen und die Studiendauer berechnet werden kann. Hierbei können sogar verschiedene erwartete Anteile an zufällig und fest rechtszensierten Beobachtungen bei der Berechnung der Stichprobenumfänge berücksichtigt werden. Die Berechnung der Fallzahlen führt ungefähr zu einer von der Nichtunterlegenheitsgrenze Δ abhängigen Aufteilung $n_S:n_N:n_P=\Delta:1:(1-\Delta)$, allerdings kann diese Aufteilung in Bezug auf den insgesamt erforderlichen Stichprobenumfang mit unserem Al-

gorithmus noch optimiert werden, mit dem Ergebnis, dass weniger Fallzahlen ausreichend sind. Auch vergleichen wir unser Verfahren mit dem vorhandenen auf der Exponentialverteilung beruhenden von [Mielke et al. \[2008a\]](#). Planen wir für beide Modelle unter der Annahme von exponentialverteilten Überlebenszeiten, so erhalten wir in etwa dieselben Fallzahlen. Anschließend untersuchen wir, wie sich das parametrische Verfahren verhält, wenn die Daten Weibull-verteilt sind mit Form-Parameter $\vartheta \neq 1$. Dort stellen wir fest, dass das parametrische Verfahren nicht robust bezüglich der Verletzung der Verteilungsannahme ist, während unsere Methode weiterhin das Niveau hält und die gewünschte Güte gut erzielt und das sogar, obwohl wir die Fallzahlen von [Mielke et al. \[2008a\]](#) zugrunde gelegt haben. Schließlich zeigen wir an zwei Beispielen, wie unser Verfahren an spezielle Frage/Problemstellungen der Praxis angepasst werden kann. Die Proportionalität der Hazardfunktionen bedeutet, dass die Therapieeffekte über die Zeit hinweg konstant sind. Zeichnet sich das neue Medikament dadurch aus, dass es z.B. nur sehr früh (oder spät) in der Zeit Vorteile gegenüber dem Standardmedikament besitzt, also nicht gleichmäßig über die Zeit hinweg, kann auch unser Modell für die gesamte Studiendauer nicht angewendet werden. Gegebenenfalls könnte die Proportionalität aber für einen gewissen Beobachtungszeitraum gelten, für den dann unser Modell wieder anwendbar wäre.

Die Fallzahlplanung und das Testverfahren zur Auswertung der Studie basieren auf asymptotischen Resultaten. Simulationen haben jedoch gezeigt, dass die asymptotischen Ergebnisse bereits für bei uns relevante Patientenzahlen herangezogen werden können.

Der allgemeiner gewählte Modellansatz hat zur Folge, dass (im Gegensatz zu dem auf der Exponentialverteilung basierendem Verfahren) die Berechnung der Fallzahlen wesentlich komplexer ist und formelmäßig nicht angegeben werden kann. Da die mit unserem Algorithmus berechneten Stichprobenumfänge nur unwesentlich von der Wahl des Form-Parameters abhängen wäre an dieser Stelle eine Vereinfachung dahingehend denkbar, dass für die Planung der Stichprobenumfänge die im parametrischen Fall vorgestellten Formeln (aus Abschnitt 7.1) verwendet und schließlich die Daten mit dem semiparametrischen Testverfahren ausgewertet werden. Da jedoch die Computer stets leistungsfähiger werden, sind komplexe Berechnungen ohne Weiteres bei akzeptablen Rechenzeiten durchführbar und stellen gegenüber direkt ausrechenbarer Formeln keine wesentliche Schwäche des semiparametrischen Verfahrens dar.

A Anhang

A.1 Herleitung der Partial-Likelihoodfunktion

Dieser Abschnitt beschäftigt sich mit dem Schätzverfahren für den Modell-Parameter des Cox-Modells. Wir zeigen, wie die Likelihoodfunktion in ein Produkt aus zwei Faktoren umgeformt wird, so dass ein Faktor alle wesentlichen Informationen enthält, der andere Faktor hingegen keine substantiellen Informationen für die Schätzung von β beisteuert. Zunächst führen wir dazu nötige Notation ein.

$T_{(l)}$: sei die l -te Ordnungsstatistik der Stichprobe $\{T_i\}$. Wir setzen $T_{(0)} := 0$ und $T_{(L+1)} := T_{n+1} := \infty$.

$Z_{(l)}$: bezeichnet die Kovariable für Person i , bei der zum Zeitpunkt $T_{(l)}$ das Ereignis eintritt.

m_l : sei für $l \in \{0, \dots, L\}$ die Anzahl der Personen, deren Zensierungszeitpunkt im Intervall $[T_{(l)}, T_{(l+1)})$ liegt.

$T_{l,j}^O \in [T_{(l)}, T_{(l+1)})$ seien für $l \in \{0, \dots, L\}$ und $j \in \{1, \dots, m_l\}$ die nach der Größe geordneten Zufallsvariablen, die die Zensierungszeitpunkte der in dem Intervall $[T_{(l)}, T_{(l+1)})$ zensierten Personen beschreiben. Person (l, j) ist diejenige, die als j -te Person im Intervall $[T_{(l)}, T_{(l+1)})$ zensiert wird. Es gilt: $T_{l,1}^O \leq \dots \leq T_{l,m_l}^O$.

$t_{l,j}^O$: sei für $l \in \{0, \dots, L\}$ und $j \in \{1, \dots, m_l\}$ die Realisierung der Zufallsvariable $T_{l,j}^O$.

Um die Likelihoodfunktion aufzustellen, führen wir Zufallsvektoren W_l und V_l ein, die angeben, bei welcher Person das Ereignis zu welcher Zeit eintritt, bzw. wann sie zensiert wird. Genauer:

W_{l+1}, V_{l+1} : Jeweils auf $Z_{(l)}$ bedingt, definiere die Zufallsvariablen W_{l+1} und V_{l+1} folgendermaßen:

W_{l+1} gebe für $l \in \{0, \dots, L\}$ an,

- bei welcher Person zum Zeitpunkt $T_{(l+1)}$ das Ereignis eintritt.

Die Realisation von W_{l+1} sei $(l+1) =: w_{l+1}$, mit $(L+1) = n+1$.

V_{l+1} gebe für $l \in \{0, \dots, L\}$ an,

- wann bei Person $(l+1)$ das Ereignis eintritt,
- wann Personen (l, j) zensiert werden $j \in \{1, \dots, m_l\}$,
- welche Personen zu den Zeitpunkten $T_{l,j}^O$, $j \in \{1, \dots, m_l\}$ zensiert werden.

Die Realisation von V_{l+1} sei der Vektor $(t_{l+1}, t_{l,1}^O, \dots, t_{l,m_l}^O, (l, 1), \dots, (l, m_l)) =: v_{l+1}$.

P_l : $P_1 := \{\emptyset\}$ und $P_l := (W_1, V_1, \dots, W_{l-1}, V_{l-1})$ für $l \in \{2, \dots, L+1\}$

P_l enthält also für alle $l \in \{2, \dots, L+1\}$:

- bei wem in $[T_{(1)}, T_{(l-1)}]$ das Ereignis eintritt,
- alle Zeitpunkte $T_{(1)}, \dots, T_{(l-1)}$, an denen ein Ereignis eintritt,
- alle Zeitpunkte in $[T_{(0)}, T_{(l-1)})$ an denen zensiert wird,
- wer in $[T_{(0)}, T_{(l-1)})$ zensiert wird.

p_l : sei die Realisation von P_l

Q_l : sei $Q_1 := V_1$ und $Q_l := (W_1, V_1, \dots, W_{l-1}, V_{l-1}, V_l)$ für $l \in \{2, \dots, L\}$

Q_l beschreibt also für alle $l \in \{2, \dots, L+1\}$:

- bei wem in $[T_{(1)}, T_{(l-1)}]$ das Ereignis eintritt,
- alle Zeitpunkte $T_{(1)}, \dots, T_{(l-1)}$, an denen ein Ereignis eintritt,
- alle Zeitpunkte in $[T_{(0)}, T_{(l-1)})$ an denen zensiert wird,
- wer in $[T_{(0)}, T_{(l-1)})$ zensiert wird.

Und zusätzlich zu P_l :

- den Zeitpunkt $T_{(l)}$,
- alle Zeitpunkte in $[T_{(l-1)}, T_{(l)})$ an denen zensiert wird,
- wer in $[T_{(l-1)}, T_{(l)})$ zensiert wird.

q_l : sei die Realisation von Q_l

Nach diesen Vorbetrachtungen geben wir nun die gemeinsame Dichte der beobachteten Werte an (Likelihoodfunktion).

$$\begin{aligned} & f_{W_1, V_1, \dots, W_L, V_L, V_{L+1}}(w_1, v_1, \dots, w_L, v_L, v_{L+1}; \beta, h_P) \\ &= \left[\prod_{l=1}^L f_{W_l|Q_l}(w_l|q_l; \beta, h_P) \right] \cdot \left[\prod_{l=1}^{L+1} f_{V_l|P_l}(v_l|p_l; \beta, h_P) \right]. \end{aligned}$$

Wir betrachten den ersten Faktor: $f_{W_l|Q_l}(w_l|q_l; \beta, h_P)$ ist die Wahrscheinlichkeit, dass bei Person (l) = w_l zum Zeitpunkt t_l das Ereignis eintritt, gegeben der zum Zeitpunkt t_l unter Risiko stehenden Personen und der Tatsache, dass zum Zeitpunkt t_l ein Ereignis eintritt. Dies kann umgeformt werden zu

$$f_{W_l|Q_l}(w_l|q_l; \beta, h_P) = \frac{e^{\beta Z_{(l)}}}{\sum_{m \in R_{t_l}} e^{\beta Z_m}}.$$

Dieser Ausdruck hängt nur noch von β und nicht mehr von h_P ab.

Eine Betrachtung des zweiten Faktors ergibt, dass dieser keine substantielle Information zur Schätzung von β beiträgt und damit weggelassen werden kann, denn falls der zweite Faktor für die Schätzung von β nicht berücksichtigt wird, wird die bedingte Wahrscheinlichkeit für

- kein Ereignis in $(T_{(l-1)}, T_{(l)})$,
- ein Ereignis in $T_{(l)}$,
- Zensierungszeiten in $[T_{(l-1)}, T_{(l)})$ und
- welche Personen in $[T_{(l-1)}, T_{(l)})$ zensiert werden

gegeben

- der in $T_{(l-1)}$ unter Risiko stehenden Personen
- dass bei Person $(l - 1)$ zum Zeitpunkt $T_{(l-1)}$ das Ereignis eintritt

für $l \in \{1, \dots, L+1\}$ nicht beachtet. Da Zensierungszeiten und Kovariablen von zensierten Personen keine oder nur wenig Informationen über den Einfluss der Kovariablen auf die Überlebenszeit liefern, können die Informationen über β , die in dem zweiten Faktor der Likelihoodfunktion enthalten sind, nicht von großer Bedeutung sein. Vor diesem Hintergrund

wird nicht mehr die volle Likelihoodfunktion zur Schätzung von β betrachtet, sondern nur noch der erste Faktor (*Partial-Likelihoodfunktion*) untersucht, nämlich

$$L(\beta) := \prod_{l=1}^L \frac{e^{\beta Z_{(l)}}}{\sum_{m \in R_{t_l}} e^{\beta Z_m}}.$$

Insbesondere kann gezeigt werden, dass der hieraus resultierende Schätzer konsistent und asymptotisch standardnormalverteilt ist (vgl. Kapitel 3.2).

A.2 Berechnung von Erwartungswerten und Ableitungen der Funktionen $S_n^{(j)}(\beta, t)$ für $j = S, N, P$

Behauptung A.1

Für $S_n^{(0)}(\beta, t)$, $S_n^{(1,S)}(\beta, t)$ und $S_n^{(1,N)}(\beta, t)$ wie in Notation 3.1 definiert gilt:

$$\begin{aligned} E[S_n^{(0)}(\beta, t)] &= P(U_1 \geq t) \left(\frac{f_S(t)}{h_P(t)} w_S + \frac{f_N(t)}{h_P(t)} w_N + \frac{f_P(t)}{h_P(t)} w_P \right), \\ E[S_n^{(1,S)}(\beta, t)] &= P(U_1 \geq t) \frac{f_S(t)}{h_P(t)} w_S, \\ E[S_n^{(1,N)}(\beta, t)] &= P(U_1 \geq t) \frac{f_N(t)}{h_P(t)} w_N. \end{aligned}$$

Beweis:

$$\begin{aligned} s^{(0)}(\beta, t) &= E[S_n^{(0)}(\beta, t)] \\ &= E[Y_1(t) e^{\beta Z_1}] \\ &= E\left[E\left[Y_1(t) \cdot e^{\beta Z_1} \mid (Z_{1S}, Z_{1N})^T\right]\right] \\ &= E\left[E\left[Y_1(t) \cdot e^{\beta Z_1} \mid (Z_{1S} = 1, Z_{1N} = 0)^T\right] \mathbf{1}_{\{Z_{1S}=1, Z_{1N}=0\}} \right. \\ &\quad \left. + E\left[Y_1(t) \cdot e^{\beta Z_1} \mid (Z_{1S} = 0, Z_{1N} = 1)^T\right] \mathbf{1}_{\{Z_{1S}=0, Z_{1N}=1\}} \right. \\ &\quad \left. + E\left[Y_1(t) \cdot e^{\beta Z_1} \mid (Z_{1S} = 0, Z_{1N} = 0)^T\right] \mathbf{1}_{\{Z_{1S}=0, Z_{1N}=0\}}\right] \\ &= E\left[P\left(Y_1(t) = 1 \mid (Z_{1S} = 1, Z_{1N} = 0)^T\right) \cdot e^{\beta S} \cdot \mathbf{1}_{\{Z_{1S}=1, Z_{1N}=0\}} \right. \\ &\quad \left. + P\left(Y_1(t) = 1 \mid (Z_{1S} = 0, Z_{1N} = 1)^T\right) \cdot e^{\beta N} \cdot \mathbf{1}_{\{Z_{1S}=0, Z_{1N}=1\}} \right. \\ &\quad \left. + P\left(Y_1(t) = 1 \mid (Z_{1S} = 0, Z_{1N} = 0)^T\right) \cdot \mathbf{1}_{\{Z_{1S}=0, Z_{1N}=0\}}\right] \\ &= e^{\beta S} \cdot P\left(Y_1(t) = 1 \mid (Z_{1S} = 1, Z_{1N} = 0)^T\right) \cdot P(Z_{1S} = 1, Z_{1N} = 0) \end{aligned}$$

$$\begin{aligned}
& + e^{\beta_N} \cdot P\left(Y_1(t) = 1 \mid (Z_{1S} = 0, Z_{1N} = 1)^T\right) \cdot P(Z_{1S} = 0, Z_{1N} = 1) \\
& + P\left(Y_1(t) = 1 \mid (Z_{1S} = 0, Z_{1N} = 0)^T\right) \cdot P(Z_{1S} = 0, Z_{1N} = 0) \\
= & e^{\beta_S} \cdot P\left(T_1 \geq t, U_1 \geq t \mid (Z_{1S} = 1, Z_{1N} = 0)^T\right) \cdot P(Z_{1S} = 1, Z_{1N} = 0) \\
& + e^{\beta_N} \cdot P\left(T_1 \geq t, U_1 \geq t \mid (Z_{1S} = 0, Z_{1N} = 1)^T\right) \cdot P(Z_{1S} = 0, Z_{1N} = 1) \\
& + P\left(T_1 \geq t, U_1 \geq t \mid (Z_{1S} = 0, Z_{1N} = 0)^T\right) \cdot P(Z_{1S} = 0, Z_{1N} = 0) \\
= & \underbrace{e^{\beta_S}}_{=\frac{h_S(t)}{h_P(t)}} \cdot \underbrace{P\left(T_1 \geq t \mid (Z_{1S} = 1, Z_{1N} = 0)^T\right)}_{=S_S(t)} \cdot \underbrace{P(Z_{1S} = 1, Z_{1N} = 0)}_{=w_S} \cdot P(U_1 \geq t) \\
& + \underbrace{e^{\beta_N}}_{=\frac{h_N(t)}{h_P(t)}} \cdot \underbrace{P\left(T_1 \geq t \mid (Z_{1S} = 0, Z_{1N} = 1)^T\right)}_{=S_N(t)} \cdot \underbrace{P(Z_{1S} = 0, Z_{1N} = 1)}_{=w_N} \cdot P(U_1 \geq t) \\
& + \underbrace{P\left(T_1 \geq t \mid (Z_{1S} = 0, Z_{1N} = 0)^T\right)}_{=S_P(t)} \cdot \underbrace{P(Z_{1S} = 0, Z_{1N} = 0)}_{=w_P} \cdot P(U_1 \geq t) \\
= & \frac{h_S(t)}{h_P(t)} \cdot S_S(t) \cdot w_S \cdot P(U_1 \geq t) + \frac{h_N(t)}{h_P(t)} \cdot S_N(t) \cdot w_N \cdot P(U_1 \geq t) \\
& + S_P(t) \cdot w_P \cdot P(U_1 \geq t) \\
= & P(U_1 \geq t) \left[\frac{f_S(t)}{h_P(t)} \cdot w_S + \frac{f_N(t)}{h_P(t)} \cdot w_N + \frac{f_P(t)}{h_P(t)} \cdot w_P \right]
\end{aligned}$$

$$\begin{aligned}
s^{(1,S)}(\beta, t) & = E \left[S_n^{(1,S)}(\beta, t) \right] \\
& = E \left[Z_{1S} Y_1(t) e^{\beta Z_1} \right] \\
& = E \left[E \left[Z_{1S} \cdot Y_1(t) \cdot e^{\beta Z_1} \mid (Z_{1S}, Z_{1N})^T \right] \right] \\
& = E \left[E \left[Z_{1S} \cdot Y_1(t) \cdot e^{\beta Z_1} \mid (Z_{1S} = 1, Z_{1N} = 0)^T \right] \mathbf{1}_{\{Z_{1S}=1, Z_{1N}=0\}} \right. \\
& \quad + E \left[Z_{1S} \cdot Y_1(t) \cdot e^{\beta Z_1} \mid (Z_{1S} = 0, Z_{1N} = 1)^T \right] \mathbf{1}_{\{Z_{1S}=0, Z_{1N}=1\}} \\
& \quad \left. + E \left[Z_{1S} \cdot Y_1(t) \cdot e^{\beta Z_1} \mid (Z_{1S} = 0, Z_{1N} = 0)^T \right] \mathbf{1}_{\{Z_{1S}=0, Z_{1N}=0\}} \right] \\
& = e^{\beta_S} \cdot P\left(T_1 \geq t, U_1 \geq t \mid (Z_{1S} = 1, Z_{1N} = 0)^T\right) \cdot P(Z_{1S} = 1, Z_{1N} = 0) \\
& = \underbrace{e^{\beta_S}}_{=\frac{h_S(t)}{h_P(t)}} \cdot \underbrace{P\left(T_1 \geq t \mid (Z_{1S} = 1, Z_{1N} = 0)^T\right)}_{=S_S(t)} \cdot \underbrace{P(Z_{1S} = 1, Z_{1N} = 0)}_{=w_S} \cdot P(U_1 \geq t) \\
& = P(U_1 \geq t) \cdot \frac{f_S(t)}{h_P(t)} \cdot w_S
\end{aligned}$$

analog dazu ergibt sich

$$s^{(1,N)}(\beta, t) = E [S_n^{(1,N)}(\beta, t)] = P(U_1 \geq t) \cdot \frac{f_N(t)}{h_P(t)} \cdot w_N$$

Behauptung A.2

Für $S_n^{(0)}(\beta, t)$, $S_n^{(1,S)}(\beta, t)$ und $S_n^{(1,N)}(\beta, t)$ wie in Notation 3.1 definiert gilt:

$$\begin{aligned} \frac{\partial \frac{S_n^{(1,S)}(\beta, t)}{S_n^{(0)}(\beta, t)}}{\partial \beta_S} &= \frac{S_n^{(1,S)}(\beta, t)}{S_n^{(0)}(\beta, t)} - \left(\frac{S_n^{(1,S)}(\beta, t)}{S_n^{(0)}(\beta, t)} \right)^2, \\ \frac{\partial \frac{S_n^{(1,S)}(\beta, t)}{S_n^{(0)}(\beta, t)}}{\partial \beta_N} &= \frac{\partial \frac{S_n^{(1,N)}(\beta, t)}{S_n^{(0)}(\beta, t)}}{\partial \beta_S} = - \frac{S_n^{(1,S)}(\beta, t) S_n^{(1,N)}(\beta, t)}{\left[S_n^{(0)}(\beta, t) \right]^2}, \\ \frac{\partial \frac{S_n^{(1,N)}(\beta, t)}{S_n^{(0)}(\beta, t)}}{\partial \beta_N} &= \frac{S_n^{(1,N)}(\beta, t)}{S_n^{(0)}(\beta, t)} - \left(\frac{S_n^{(1,N)}(\beta, t)}{S_n^{(0)}(\beta, t)} \right)^2. \end{aligned}$$

Beweis:

$$\begin{aligned}
\frac{\partial \frac{S_n^{(1,S)}(\beta,t)}{S_n^{(0)}(\beta,t)}}{\partial \beta_S} &= \frac{\partial \frac{1}{n} \sum_{i=1}^n Z_{iS} e^{\beta Z_i Y_i(t)}}{\partial \beta_S \frac{1}{n} \sum_{i=1}^n e^{\beta Z_i Y_i(t)}} \\
&= \frac{\left(\frac{1}{n} \sum_{i=1}^n Z_{iS}^2 e^{\beta Z_i Y_i(t)} \right) \left(\frac{1}{n} \sum_{i=1}^n e^{\beta Z_i Y_i(t)} \right) - \left(\frac{1}{n} \sum_{i=1}^n Z_{iS} e^{\beta Z_i Y_i(t)} \right) \left(\frac{1}{n} \sum_{i=1}^n Z_{iS} e^{\beta Z_i Y_i(t)} \right)}{\left(\frac{1}{n} \sum_{i=1}^n e^{\beta Z_i Y_i(t)} \right)^2} \\
&= \frac{\frac{1}{n} \sum_{i=1}^n Z_{iS} e^{\beta Z_i Y_i(t)}}{\frac{1}{n} \sum_{i=1}^n e^{\beta Z_i Y_i(t)}} - \left(\frac{\frac{1}{n} \sum_{i=1}^n Z_{iS} e^{\beta Z_i Y_i(t)}}{\frac{1}{n} \sum_{i=1}^n e^{\beta Z_i Y_i(t)}} \right)^2 = \frac{S_n^{(1,S)}(\beta,t)}{S_n^{(0)}(\beta,t)} - \left(\frac{S_n^{(1,S)}(\beta,t)}{S_n^{(0)}(\beta,t)} \right)^2
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \frac{S_n^{(1,S)}(\beta,t)}{S_n^{(0)}(\beta,t)}}{\partial \beta_N} &= \frac{\partial \frac{1}{n} \sum_{i=1}^n Z_{iS} e^{\beta Z_i Y_i(t)}}{\partial \beta_N \frac{1}{n} \sum_{i=1}^n e^{\beta Z_i Y_i(t)}} \\
&= \frac{\left(\frac{1}{n} \sum_{i=1}^n Z_{iS} Z_{iN} e^{\beta Z_i Y_i(t)} \right) \left(\frac{1}{n} \sum_{i=1}^n e^{\beta Z_i Y_i(t)} \right) - \left(\frac{1}{n} \sum_{i=1}^n Z_{iS} e^{\beta Z_i Y_i(t)} \right) \left(\frac{1}{n} \sum_{i=1}^n Z_{iN} e^{\beta Z_i Y_i(t)} \right)}{\left(\frac{1}{n} \sum_{i=1}^n e^{\beta Z_i Y_i(t)} \right)^2} \\
&= - \frac{\left(\frac{1}{n} \sum_{i=1}^n Z_{iS} e^{\beta Z_i Y_i(t)} \right) \left(\frac{1}{n} \sum_{i=1}^n Z_{iN} e^{\beta Z_i Y_i(t)} \right)}{\left(\frac{1}{n} \sum_{i=1}^n e^{\beta Z_i Y_i(t)} \right)^2} = \frac{\partial \frac{S_n^{(1,N)}(\beta,t)}{S_n^{(0)}(\beta,t)}}{\partial \beta_S} = - \frac{S_n^{(1,S)}(\beta,t) S_n^{(1,N)}(\beta,t)}{\left[S_n^{(0)}(\beta,t) \right]^2}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \frac{S_n^{(1,N)}(\beta,t)}{S_n^{(0)}(\beta,t)}}{\partial \beta_N} &= \frac{\partial \frac{1}{n} \sum_{i=1}^n Z_{iN} e^{\beta Z_i Y_i(t)}}{\partial \beta_N \frac{1}{n} \sum_{i=1}^n e^{\beta Z_i Y_i(t)}} \\
&= \frac{\left(\frac{1}{n} \sum_{i=1}^n Z_{iN}^2 e^{\beta Z_i Y_i(t)} \right) \left(\frac{1}{n} \sum_{i=1}^n e^{\beta Z_i Y_i(t)} \right) - \left(\frac{1}{n} \sum_{i=1}^n Z_{iN} e^{\beta Z_i Y_i(t)} \right) \left(\frac{1}{n} \sum_{i=1}^n Z_{iN} e^{\beta Z_i Y_i(t)} \right)}{\left(\frac{1}{n} \sum_{i=1}^n e^{\beta Z_i Y_i(t)} \right)^2} \\
&= \frac{\frac{1}{n} \sum_{i=1}^n Z_{iN} e^{\beta Z_i Y_i(t)}}{\frac{1}{n} \sum_{i=1}^n e^{\beta Z_i Y_i(t)}} - \left(\frac{\frac{1}{n} \sum_{i=1}^n Z_{iN} e^{\beta Z_i Y_i(t)}}{\frac{1}{n} \sum_{i=1}^n e^{\beta Z_i Y_i(t)}} \right)^2 = \frac{S_n^{(1,N)}(\beta,t)}{S_n^{(0)}(\beta,t)} - \left(\frac{S_n^{(1,N)}(\beta,t)}{S_n^{(0)}(\beta,t)} \right)^2
\end{aligned}$$

A.3

Satz A.3

Eine Hazardfunktion ist genau dann konstant, wenn eine Exponentialverteilung zugrunde liegt.

Beweis:

Sei $c_1 \in \mathbb{R}$ eine Konstante, dann gilt

$$\begin{aligned} c_1 &\stackrel{\text{Bem. 2.2}}{=} \frac{f(t)}{S(t)} = \frac{-dS(t)/dt}{S(t)} = -\frac{d \log(S(t))}{dt} && \text{für alle } t \in \mathbb{B} = [0, b] \\ \Leftrightarrow \log(S(t)) &= -c_1 t + c_2 && \text{für alle } t \in \mathbb{B} \text{ mit } c_2 \in \mathbb{R} \\ \Leftrightarrow S(t) &= e^{-c_1 t + c_2} && \text{für alle } t \in \mathbb{B}. \end{aligned}$$

Da $S(0) = 1 = e^{c_2}$, ist $S(t) = e^{-c_1 t}$ und $F(t) = 1 - e^{-c_1 t}$, welche die Verteilungsfunktion der Exponentialverteilung ist. □

Literaturverzeichnis

- O. Aalen. Nonparametric inference for a family of counting processes. *The Annals of Statistics*, 6:701–726, 1978.
- P. K. Andersen and R. D. Gill. Cox’s regression model for counting processes: a large sample study. *The Annals of Statistics*, 10:1100–1120, 1982.
- P. K. Andersen, O. Borgan, R. D. Gill, and N. Keiding. *Statistical Methods Based on Counting Processes*. Springer Series in Statistics. Springer, New York, 1993.
- M. V. P. Bernardo and D. P. Harrington. Sample size calculations for the two-sample problem using the multiplicative intensity model. *Statistics in Medicine*, 20:557–579, 2001.
- D. Brown, P. Volkers, and S. Day. An introductory note to CHMP guidelines: Choice of the non-inferiority margin and data monitoring committees. *Statistics in Medicine*, 25:1623–1627, 2006.
- G. Casella and R. L. Berger. *Statistical Inference*. Duxbury Press, 1990.
- CHMP. (Committee for Medicinal Products for Human Use). Guideline on the choice of the non-inferiority margin. EMEA/CPMP/EWP/2158/99, 2005.
- D. Collett. *Modelling Survival Data in Medical Research*. Texts in Statistical Science. Chapman & Hall, London, 1996.
- A. R. Conn, K. Scheinberg, and L. N. Vicente. *Introduction to derivative-free optimization*. MPS/SIAM Series on Optimization 8. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM), New York, 2009.

- D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society, Series B*, 34:187–220, 1972.
- D. R. Cox. Partial likelihood. *Biometrika*, 62:269–276, 1975.
- CPMP. (Committee for Proprietary Medicinal Products). Points to consider on adjustment for baseline covariates. www.ema.europa.eu/pdfs/human/ewp/286399en.pdf, CPMP/EWP/2863/99, 2003.
- FDA. (Food and Drug Administration). Guidance for industry on non-inferiority clinical trials. 2010.
- T. S. Ferguson. *A Course in Large Sample Theory*. Texts in Statistical Science. Chapman & Hall, London, 1996.
- T. R. Fleming and D. P. Harrington. *Counting Processes and Survival Analysis*. Wiley series in Probability and Statistics. Wiley, New York, 2005.
- M. Hasler, R. Vonk, and L. A. Hothorn. Assessing non-inferiority of a new treatment in a three-arm trial in the presence of heteroscedasticity. *Statistics in Medicine*, 27:490–503, 2008.
- D. Hauschke and I. Pigeot. Establishing efficacy of a new experimental treatment in the ‘gold standard’ design. *Biometrical Journal*, 47:782–786, 2005.
- H. M. J. Hung, S.-J. Wang, and R. O’Neil. A regulatory perspective on choice of margin and statistical inference issue in non-inferiority trials. *Biometrical Journal*, 47:28–36, 2005.
- ICH. Harmonised tripartite guideline. Statistical principles for clinical trials, ICH topic E9. www.ema.europa.eu/pdfs/human/ich/036396en.pdf, 1998.
- M. Kieser and T. Friede. Planning and analysis of three-arm non-inferiority trials with binary endpoints. *Statistics in Medicine*, 26:253–273, 2007.
- J. P. Klein and M. L. Moeschberger. *Survival Analysis - Techniques for censored and Truncated Data, Second Edition*. Statistics for Biology and Health. Springer, New York, 2005.
- D. G. Kleinbaum. *Survival Analysis*. Statistics in the health sciences. Springer, Berlin, 1999.

-
- A. Koch. Discussion on “establishing efficacy of a new experimental treatment in the ‘gold standard’ design”. *Biometrical Journal*, 47:792–793, 2005.
- A. Koch and J. Röhmel. Hypothesis testing in the “gold standard” design for proving the efficacy of an experimental treatment relative to placebo and a reference. *Journal of Biopharmaceutical Statistics*, 414:315–325, 2004.
- S. Lange and G. Freitag. Choice of delta: requirements and reality - results of a systematic review. *Biometrical Journal*, 47:12–27, 2005.
- E. Marubini and M. G. Valsecchi. *Analysing Survival Data from Clinical Trials and Observational Studies*. Statistics in practice. Wiley, Chichester, 1995.
- M. Mielke and A. Munk. The assessment and planning of non-inferiority trials for retention of effect hypotheses - towards a general approach. [arXiv:0912.4169v1 \[stat.ME\]](https://arxiv.org/abs/0912.4169v1).
- M. Mielke, A. Munk, and A. Schacht. Planning and assessing non-inferiority in a gold standard design with censored, exponentially distributed endpoints. *Statistics in Medicine*, 27:5093–5110, 2008a.
- M. Mielke, A. Munk, and A. Schacht. Technical report on “the assessment of non-inferiority in a gold standard design with censored, exponentially distributed endpoints”, 2008b. www.stochastik.math.uni-goettingen.de/preprints/mielke2008_TechnicalReport.pdf.
- S. A. Montgomery. Clinically relevant effect sizes in depression. *European Neuropsychopharmacology*, 4:283–284, 1994.
- I. Pigeot, J. Schäfer, J. Röhmel, and D. Hauschke. Assessing non-inferiority of a new treatment in a three-arm clinical trial including a placebo. *Statistics in Medicine*, 22:883–899, 2003.
- J. Röhmel, D. Hauschke, A. Koch, and I. Pigeot. Biometrische Verfahren zum Wirksamkeitsnachweis im Zulassungsverfahren. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, 48:562–571, 2005.
- M. Rothmann, N. Li, G. Chen, G. Y. H. Chi, R. Temple, and H.-H. Tsou. Design and analysis of non-inferiority mortality trials in oncology. *Statistics in Medicine*, 22:239–264, 2001.

- K. F. Schulz and D. A. Grimes. Generation of allocation sequences in randomised trials: chance, not choice. *The Lancet*, 359:515–519, 2002.
- S. M. Snapinn. Alternatives for discounting in the analysis of noninferiority trials. *Journal of Biopharmaceutical Statistics*, 14:263–273, 2004.
- R. Temple and S. S. Ellenberg. Placebo-controlled trials and active-controlled trials in the evaluation of new treatments part 1: Ethical and scientific issues. *Annals of Internal Medicine*, 133:455–463, 2000.
- T. M. Therneau and P. M. Grambsch. *Modeling Survival Data: Extending the Cox Model*. Statistics for Biology and Health. Springer, New York, 2001.
- A. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, New York, 1998.
- WMA. (World Medical Association) Declaration of Helsinki - ethical principles for medical research involving human subjects. www.wma.net/en/30publications/10policies/b3/, 2008.