

# Ruprecht-Karls-Universität Heidelberg



Masterarbeit zur Erlangung des akademischen Grades  
„Master of Science in Medical Biometry/Biostatistics“

>>Über den Einfluss unbalanzierter  
Randomisierungsverhältnisse und Gruppengrößen  
sowie weiterer Designparameter auf die Genauigkeit  
verschiedener Schätzmethode der Hazard Ratio für  
systematische Übersichtsarbeiten und Metaanalysen  
– eine Simulationsstudie und  
Anwendungsempfehlung<<

vorgelegt von

>>Dr. med. Benjamin Thomas<<

>>2023<<

**>>Über den Einfluss unbalanzierter  
Randomisierungsverhältnisse und Gruppengrößen  
sowie weiterer Designparameter auf die Genauigkeit  
verschiedener Schätzmethoden der Hazard Ratio für  
systematische Übersichtsarbeiten und Metaanalysen  
– eine Simulationsstudie und Anwendungsempfehlung<<**

vorgelegt von: Dr. med. Benjamin Thomas

Studiengang: MSc Medical Biometry/Biostatistics

Betreuer und Erstgutachter: Dr. phil. Marietta Kirchner  
Universität Heidelberg

Zweitgutachter: Prof. Dr. Peter Martus  
Universität Tübingen

Abgabedatum: 09.03.2023

## **Erklärung**

Hiermit erkläre ich, dass ich die vorliegende Masterarbeit selbständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe.

Die Stellen der Masterarbeit, die anderen Quellen im Wortlaut oder dem Sinn nach entnommen wurden, sind durch Angaben der Herkunft kenntlich gemacht. Dies gilt auch für Zeichnungen, Skizzen, bildliche Darstellungen sowie für Quellen aus dem Internet.

Mannheim, den 08.03.2023

---

Benjamin Thomas

## Verwendete Notationen

Alle verwendeten mathematischen Formeln werden gemäß allgemeingültigen mathematischen Notationen eingeführt und fortlaufend in runden Klammern () referenziert.

R-Code bzw. Code-Ausschnitte von besonderer Relevanz werden im Fließtext gesondert formatiert:

```
Code # Erläuterung
```

Der verwendeten Programmiersprache R entsprechend, sind alle Erläuterungen im Code in englischer Sprache verfasst.

Da die Veröffentlichung der Inhalte der vorliegenden Masterarbeit in einem internationalen Fachjournal in englischer Sprache geplant ist, wurden alle Beschriftungen innerhalb von Abbildungen und Tabellen ebenso in englischer Sprache verfasst.

Um Verwirrungen zu vermeiden, wurden Dezimalstellen auch im deutschen Fließtext der verwendeten Programmiersprache R entsprechend mit Punkten anstatt Kommata gekennzeichnet.

## Glossar verwendeter Abkürzungen und Begrifflichkeiten

Abb. = Abbildung

ADEMP = Akronym, steht für Aims, Data-generating mechanisms, Estimands, Methods und Performance measures

$\alpha$  = (berichtetes) Signifikanzniveau

APA = American Psychological Association

$C_{\text{exp}}$  = Exponentialverteilung auftretender (Rechts-) Zensierungen

$C_{\text{unif}}$  = Gleichverteilung auftretender (Rechts-) Zensierungen

CI = Konfidenzintervall (englisch: *confidence interval*)

DGM = Data-generating mechanisms

et al. = et aliae bzw. et alii

E = Erwartungswert

Exp = Exponentialverteilung bzw. exponentialverteilt

$h(t)$  = Funktion der Ausfallrate, Ereignisrate (englisch: *hazard function*)

$h_0(t)$  = Funktion der Ausfallrate, Ereignisrate der Kontrollgruppe

$h_1(t)$  = Funktion der Ausfallrate, Ereignisrate der Experimentalgruppe

$H(t)$  = kumulative Hazardfunktion (englisch: *cumulative hazard function*)

HR bzw.  $HR_i$  = Hazard Ratio

$HR_{\text{obs}}$  = im Rahmen der Cox-Regression beobachtetes Hazard Ratio

$HR_{\text{true}}$  = wahres Hazard Ratio

i.d.R. = in der Regel

inkl. = inklusive

$\lambda$  = Ausfallrate, Ereignisrate der Kontrollgruppe (englisch: *baseline hazard*)

$\ln$  = natürlicher Logarithmus, gleichbedeutend mit  $\log$

(die Annotationen  $\ln$  und  $\log$  werden im Folgenden nicht differenziert)

$\log$  = natürlicher Logarithmus, gleichbedeutend mit  $\ln$

(die Annotationen  $\ln$  und  $\log$  werden im Folgenden nicht differenziert)

$\log \text{ var}$  = Log-Rang-Varianz (englisch: *logrank variance*)

MCSE = Monte Carlo Standardfehler (englisch: *Monte Carlo Standard Error*)

MSE = mittlerer quadrierter Fehler

$n_{\text{obs}}$  = Gesamtfallzahl einer Studie

$n_{\text{sim}}$  = Anzahl an Simulationsdurchläufen

o.g. = oben genannt

$p$  = (berichteter) P-Wert

$\Phi$  = Verteilungsfunktion der Standardnormalverteilung (griechisch: *Phi*)

$r$  = Allokationsrate

RMSE = Wurzel des mittleren quadrierten Fehlers

s. = siehe

$S(t)$  = Überlebensfunktion (englisch: *survival function*)

sog. = sogenannte

$t$  = Zeitpunkt  $t$

Tab. = Tabelle

$\theta$  = unbekannter Parameter (griechisch: *Theta*)

$\hat{\theta}$  = Schätzer für den unbekannter Parameter

$\bar{\theta}$  = arithmetischer Mittelwert der Schätzer für den unbekannter Parameter

Unif = Gleichverteilung bzw. gleichverteilt

$V_i$  = Varianz der Studie  $i$  im Rahmen einer Metanalyse

var = Varianz

vgl. = vergleiche

$w_i$  = Gewicht der Studie  $i$  im Rahmen einer Metanalyse

z.B. = zum Beispiel



## Inhaltsverzeichnis

<b>Titel</b>	<b>2</b>
<b>Erklärung</b>	<b>3</b>
<b>Verwendete Notationen</b>	<b>4</b>
<b>Glossar verwendeter Abkürzungen und Begrifflichkeiten</b>	<b>5</b>
<b>Inhaltsverzeichnis</b>	<b>8</b>
<b>Danksagung</b>	<b>10</b>
<b>Einleitung</b>	<b>11</b>
Einführung in die Überlebenszeitanalyse	11
Einführung in die Metaanalyse	13
Die Inverse Varianz-Methode	14
Metaanalyse von Überlebenszeitstudien	14
Aktueller Stand der Forschung	16
Wissenschaftlich-biometrische Fragestellung der vorliegenden Arbeit	17
Biometrisch-biostatistische Bedeutung von Simulationsstudien	17
<b>Material und Methoden</b>	<b>19</b>
Szenarien der vorgelegten Simulationsstudie	19
Aims: motivierende Fragestellung der vorliegenden Simulationsstudie	20
Data-generating mechanisms: Daten-generierende Modelle in der vorliegenden Simulationsstudie	21
1) Binäre Kovariable "Treatment"	21
2) Univariat exponentialverteilte Kovariable "Survivaltime"	21
3) Univariat exponential-/uniformverteilte Kovariable "Censoring"	22
4) Univariat exponentialverteilte Kovariable "Observed" in Abhängigkeit von der Zensierungsrate	22
Estimands: zu schätzende Zielgrößen der vorliegenden Simulationsstudie	22
Methods: beobachtete log HRs und Varianzen	23
Cox-Regression	23
Methoden 1 bis 9 nach Tierney	24
Berechnungsmethoden nach Tierney umgesetzt in RStudio	27
Performance measures 1: Erwartungstreue und Verzerrtheit der Punktschätzer für log(HR) in der vorliegenden Simulationsstudie	29
Verzerrtheit (Bias)	29
Monte Carlo Standardfehler des Bias (MCSE)	30
Varianz der Punktschätzer	30
Der mittlere quadrierte Fehler (MSE)	30
Die Wurzel des mittleren quadrierten Fehlers (RMSE)	31
calculate_performances()	31
Performance measures 2: Punktschätzer für var(log(HR)) und jeweiliger prozentualer Gewichtanteil in der vorliegenden Simulationsstudie	33
calculate_variances()	33
Relative Studiengewichte	34



<b>Ergebnisse</b> -----	<b>36 -</b>
<b>Erzeugte Ergebnisdatenmatrix</b> -----	<b>36 -</b>
Deskriptive Statistik der Ergebnisdatenmatrix "GRID" -----	36 -
<b>Performance measures der log(HR)-Schätzmethoden pro Hazard Ratio</b> -----	<b>38 -</b>
Gepoolte $\log(HR)$ -Schätzer pro Hazard Ratio -----	40 -
Gepoolter Bias der $\log(HR)$ -Schätzer pro Hazard Ratio-----	41 -
<b>Gepoolte log(HR)-Schätzer und gepoolter Bias pro Hazard Ratio unter faktorieller Variation der Designparameter „Allocation Ratio“, „Baseline Hazard“, „Censoring Distribution“ und „Trial Size“</b> -----	<b>42 -</b>
Bias und Verteilung der Punktschätzer unter faktorieller Variation der Designparameter „Allocation Ratio“, „Baseline Hazard“, „Censoring Distribution“ und „Trial Size“ bei Hazard Ratio von 0.1:-----	43 -
Bias und Verteilung der Punktschätzer unter faktorieller Variation der Designparameter „Allocation Ratio“, „Baseline Hazard“, „Censoring Distribution“ und „Trial Size“ bei Hazard Ratio von 0.3:-----	45 -
Bias und Verteilung der Punktschätzer unter faktorieller Variation der Designparameter „Allocation Ratio“, „Baseline Hazard“, „Censoring Distribution“ und „Trial Size“ bei Hazard Ratio von 0.5:-----	47 -
Bias und Verteilung der Punktschätzer unter faktorieller Variation der Designparameter „Allocation Ratio“, „Baseline Hazard“, „Censoring Distribution“ und „Trial Size“ bei Hazard Ratio von 0.7:-----	49 -
Bias und Verteilung der Punktschätzer unter faktorieller Variation der Designparameter „Allocation Ratio“, „Baseline Hazard“, „Censoring Distribution“ und „Trial Size“ bei Hazard Ratio von 0.9:-----	51 -
<b>Performance measures der var(log(HR))-Punktschätzer</b> -----	<b>53 -</b>
<b>Relative Studiengewichte basierend auf verschiedenen var(log(HR))-Schätzern unter faktorieller Variation aller Designparameter</b> -----	<b>53 -</b>
<b>Diskussion und Fazit</b> -----	<b>56 -</b>
<b>Rekapitulation der Ausgangslage und Motivation der Fragestellung</b> -----	<b>56 -</b>
<b>Anhand bisher verfügbarer Literatur erwartbarer Ergebnisanteil</b> -----	<b>57 -</b>
Bias der $\log(HR_{obs})$ -Schätzer im Cox-Modell-----	57 -
Bias der approximativen $\log(HR)$ - und $\text{var}(\log(HR))$ -Schätzer-----	57 -
<b>Zusammenfassung der Ergebnisse der vorliegenden Simulationsstudie</b> -----	<b>57 -</b>
Bias und MCSE der approximativen $\log(HR)$ -Schätzer -----	58 -
<b>Interpretation und Bewertung der Ergebnisse der vorliegenden Simulationsstudie</b> -----	<b>58 -</b>
Beobachteter Bias der $\log(HR)$ -Schätzer der Cox-Regression-----	58 -
Beobachteter Bias der $\log(HR)$ -Schätzer der Methoden 1 bis 9-----	59 -
Approximative $\text{var}(\log(HR))$ -Schätzer und relative Gewichte der Methoden 1 bis 9-----	60 -
<b>Anwendungsempfehlungen für die Praxis</b> -----	<b>61 -</b>
<b>Limitationen der vorliegenden Simulationsstudie</b> -----	<b>62 -</b>
Verwendung mittels Pseudozufallszahlen simulierter Datensätze -----	62 -
Verwendung exponentialverteilter Überlebenszeitdaten, proportionaler Hazards und exponential-/uniformverteilter Zensierungen -----	63 -
<b>Ausblick und Ansatzpunkte für weiterführende Untersuchungen</b> -----	<b>64 -</b>
<b>Abbildungsverzeichnis</b> -----	<b>65 -</b>
<b>Tabellenverzeichnis</b> -----	<b>66 -</b>
<b>Literaturverzeichnis</b> -----	<b>67 -</b>
<b>Verzeichnis verwendeter R-Pakete</b> -----	<b>71 -</b>
<b>Anhang</b> -----	<b>72 -</b>

## **Danksagung**

Auf diesem Wege möchte ich bei all denjenigen Unterstützerinnen und Unterstützern bedanken, die mich während der letzten drei Jahre und insbesondere der letzten sechs Monate auf dem Weg zur Erlangung des Master of Science in medizinischer Biometrie und Biostatistik begleitet haben.

Allen voran gebührt mein Dank Frau Dr. Marietta Kirchner, die meinen Initiativvorschlag für das Thema dieser Masterarbeit von Anfang an unterstützte. Für ihre kontinuierliche Betreuung, ihre zahlreichen hilfreichen Anregungen und ihr konstruktives Feedback bei der Erstellung dieser Arbeit und nicht zuletzt für die Begutachtung der Arbeit möchte ich mich herzlich bedanken. Darüber hinaus möchte ich mich insbesondere auch für die hervorragende Organisation des gesamten Studienganges gerade in den schwierigen Zeiten der Corona-Pandemie bedanken. An dieser Stelle möchte ich mich auch bei Frau Andrea Wendel bedanken, die den Studierenden immer mit Rat und Tat zur Seite stand. Mein Dank gilt außerdem Herrn Prof. Dr. Martus, der sich bereit erklärt hat, die vorliegende Arbeit als Zweitgutachter zu bewerten.

Herzlich bedanken möchte ich mich auch bei Herrn Prof. Dr. Kneser, Direktor der Klinik für Hand-, Plastische-, Rekonstruktive und Mikrochirurgie an der BG Unfallklinik Ludwigshafen, meinem Chef, der mich in all meinen Bestrebungen über den klinischen Tellerrand zu blicken immer unterstützt hat und meine Teilnahme an diesem berufsbegleitenden Studiengang durch seine Unterstützung erst möglich gemacht hat. Für die hohe Flexibilität und uneingeschränkte Unterstützung in der Umsetzung der Personalplanung zur Ermöglichung der zahlreichen Studientage möchte ich mich bei Herrn PD Dr. Gazyakan, Herrn PD Dr. Bigdeli und Frau Prof. Harhaus-Wähner bedanken.

Zuletzt möchte ich einen besonderen Dank an meine Partnerin und meine Eltern richten, die mich durch die schöne, aber auch anstrengende Zeit des Studiums und der Erstellung der vorliegenden Arbeit begleitet haben und auch in schwierigen Zeiten immer an meiner Seite waren.

## Einleitung

### *Einführung in die Überlebenszeitanalyse*

Sogenannte *time-to-event* Daten stellen die Grundlage der Überlebenszeitanalyse dar. Typischerweise wird in diesem Kontext die Wirksamkeit einer Intervention anhand des Vergleichs der Zeiträume bis zum Auftreten eines definierten Ereignisses innerhalb eines Beobachtungsintervalles zwischen Kontroll- und Experimentalgruppe beurteilt. Mögliche Anwendungsbereiche erstrecken sich von der Ableitung von mittleren Erwartungszeiträumen über die Bewertung von therapeutischen Wirksamkeiten bis hin zur Vorhersage von entitätsspezifischen Überlebenszeitprognosen.<sup>1</sup> Die Struktur der zugrundeliegenden Daten weist dabei folgende spezifische Gemeinsamkeiten auf:

1. Dem eigentlichen Beobachtungszeitraum (sog. „follow-up time“) ist im Rahmen prospektiver Studiendesigns eine Rekrutierungsphase vorgeschaltet (sog. „accrual time“).
2. Das Zielereignis muss nicht zwangsläufig für alle Probanden innerhalb des Beobachtungszeitraumes auftreten (sog. „Zensierungen“).
3. Das Zielereignis kann auch bedingt durch die Exposition gegenüber einem alternativen Risiko aufgetreten sein (sog. „competing risk“).
4. Die Verteilungen von Überlebenszeiten folgen selten einer Normalverteilung, sondern weisen i.d.R. eine deutliche Rechtsschiefe auf.

Den vorgenannten und weiteren Aspekten Rechnung tragend wurden spezielle Auswertungs- und Analyseansätze entwickelt, die im Wesentlichen auf dem Vergleich zweier in Zusammenhang stehender Wahrscheinlichkeitsverteilungen beruhen – der *Survivalfunktion* und der *Hazardfunktion*.<sup>2</sup> Die *Überlebensfunktion* beschreibt dabei die Wahrscheinlichkeit, einen Zeitpunkt  $t$  ereignisfrei zu überleben und ist definiert als  $S(t)$ . Der negative natürliche Logarithmus der *Überlebensfunktion* –  $\log(S(t))$  beschreibt das Integral vom Zeitpunkt 0 bis zum Zeitpunkt  $t$  über die *Ausfallraten*  $\int_0^t h(t)$  als *kumulative Hazardfunktion*  $H(t)$ .

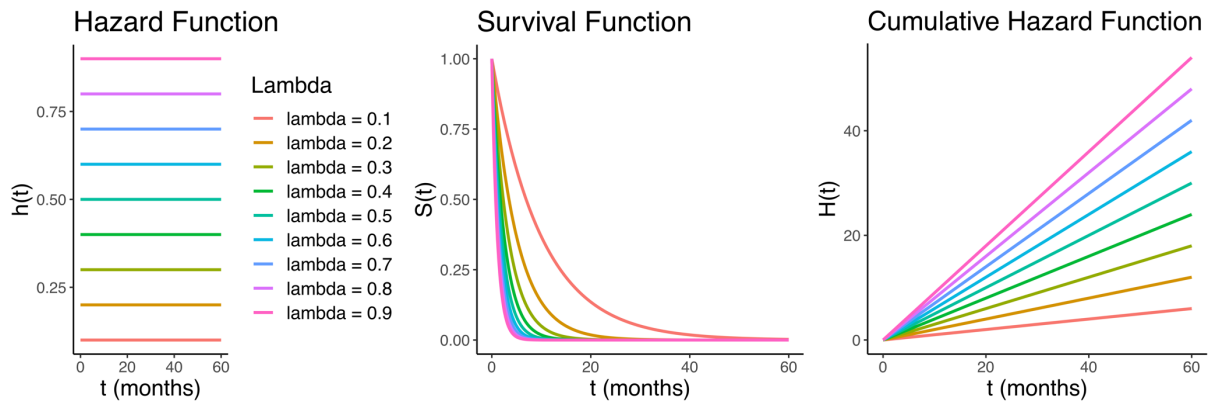


Abb. 1: Zusammenhang zwischen  $h(t)$ ,  $S(t)$  und  $H(t)$ . Konstante Ereignisraten bedingen exponentialverteilte Überlebenskurven und monoton ansteigende kumulative Inzidenzdichten.

Das Hazard Ratio  $HR = \frac{h_1(t)}{h_0(t)}$  ermöglicht schließlich den Vergleich zweier Gruppen, indem man deren Hazards ins Verhältnis setzt.

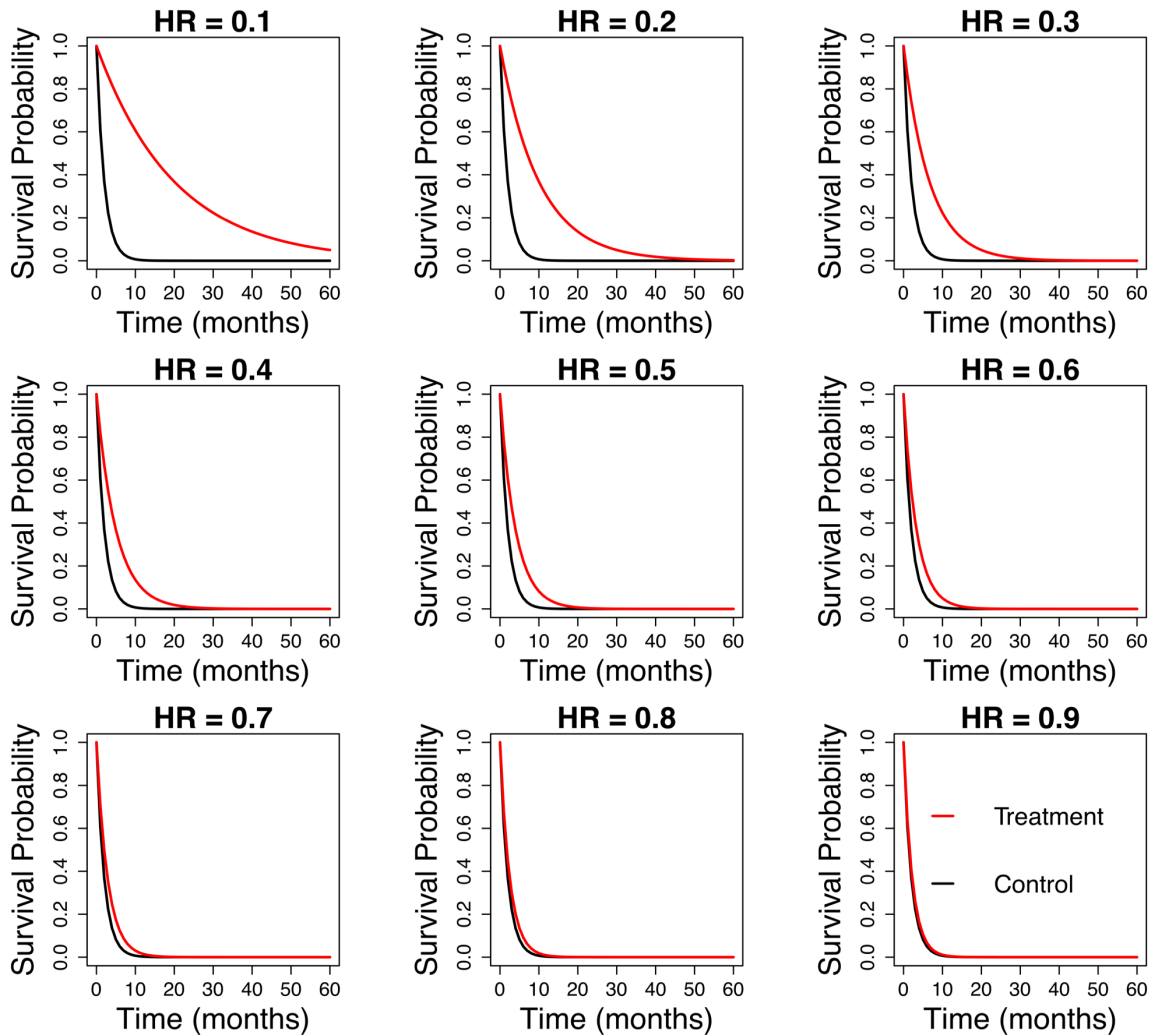


Abb. 2: Hazard Ratios quantifizieren die Verhältnisse von Ausfallraten unter Therapie versus Kontrolle im selben Zeitintervall. Ein Wert von 1 entspricht keinem Unterschied. Ein Wert von 0.1 bedeutet eine 90%-ige Wahrscheinlichkeitsreduktion des Ereigniseintritts unter Therapie.

### *Einführung in die Metaanalyse*

Um Entscheidungen über Diagnostik, Behandlung und Versorgung auf einer soliden Evidenzgrundlage zu treffen, ist die systematische Analyse verfügbarer biomedizinischer Daten unverzichtbar.<sup>3</sup> Um eine präzisere Schätzung wissenschaftlich untersuchter Behandlungseffekt zu erlangen, als dies durch die individuelle Analyse einzelner Studien möglich wäre, werden Daten aus mehreren Studien mit gemeinsamen Fragestellungen daher im Rahmen von Metaanalysen kombiniert.<sup>4</sup> Die Metaanalyse ist damit ein wichtiger Teil des Evidenz-basierten Denkens und stellt die Methode der Wahl zur Überprüfung einzelner und Synthese mehrere Ergebnisse publizierter Studien zu einem bestimmten Thema oder einer bestimmten Fragestellung dar.<sup>5</sup> Durch die Verwendung eigens zu diesem Zweck entwickelter biostatistischer Methoden kann im Rahmen einer Metaanalyse eine präzisere Schätzung zugrundeliegender Behandlungseffekte erreicht werden, indem Verzerrungen einzelner Studien aufgedeckt und ausgeglichen werden. Die Methodik einer Metaanalyse umfasst fünf Hauptbestandteile:<sup>6</sup>

1. Die Identifikation aller relevanten Studien.
2. Die Überprüfung der Qualität aller identifizierten Studien.
3. Die Synthese aller Daten aus den eingeschlossenen Studien.
4. Die Überprüfung der Varianz der Effektschätzer der einzelnen Studien.
5. Die Visualisierung und Interpretation der Ergebnisse.

Die Varianzanalysetechniken bei der Datensynthese können im Wesentlichen in zwei Ansätze unterteilt werden, den *Modellen mit festen Effekten* und den *Modellen mit zufälligen Effekten*.<sup>7</sup> Festeffekte-Modelle betrachten den Effekt einer Intervention als konstant und gleich für alle Studien in einer Metaanalyse und berechnen den durchschnittlichen Effekt über alle Studien. Zufällige Effekte-Modelle hingegen berücksichtigen die Tatsache, dass der Effekt einer Intervention von Studie zu Studie variieren kann. Diese Modelle berechnen für jede Studie einen individuellen Effektschätzer und berechnen im Nachgang einen durchschnittlichen Effekt, der die vorgenannte Variabilität der studienspezifischen Effekte berücksichtigt.

Das Ergebnis einer Metaanalyse wird dabei i.d.R. in Form von statistischen Kenngrößen, bestehend aus gemittelten Effektmaßen mit dazu gehörigen Konfidenzintervallen, präsentiert. Um Unterschieden in der Qualität eingeschlossener und analysierter Studien dabei entsprechend Rechnung zu tragen, werden diese

unterschiedlich gewichtet, um den Einfluss individueller Studien auf das Gesamtergebnis der Metaanalyse zu berücksichtigen und ggf. zu korrigieren.<sup>8</sup> In diesem Rahmen ist die Studiengröße neben der Qualität des Studiendesigns ein wichtiger Parameter, denn es besteht ein gegenläufiger Zusammenhang zwischen der Anzahl der Studienteilnehmer und der Unsicherheit des studienspezifischen Effektschätzers. Generell gilt, dass größere Studien eine geringere Varianz und somit eine geringere Fehlerrate bei der Effektschätzung aufweisen, wohingegen kleinere Studien zu einer höheren Varianz und einer höheren Fehlerrate führen.

### *Die Inverse Varianz-Methode*

Zur Gewichtung von Studien in einer Metaanalyse kommen i.d.R. Vertreter der inversen Varianz-Methode zum Einsatz.<sup>8</sup> Die grundlegende Idee hinter dieser Methode ist, dass Studien mit einer kleineren Varianz und entsprechend geringerer Unsicherheit der Effektschätzer ein höheres Gewicht und damit mehr Einfluss bei der Berechnung des gepoolten Effektschätzers erhalten sollten als Studien mit einer größeren Varianz und höherer Unsicherheit. Dies wird praktisch erreicht, indem die Varianz jeder Studie berechnet wird und dann die inverse Varianz dieser Studie als prozentualer Anteil der inversen Gesamtvarianz verwendet wird. Auf diese Weise erhält jede Studie ein höheres Gewicht, wenn ihre Varianz gering ist und ein geringeres Gewicht, wenn ihre Varianz hoch ist. In ihrer ursprünglichsten Form ist die inverse Varianz-Methode einfach zu berechnen und einfach zu interpretieren.

$$w_{fixed} = \frac{1}{V} \quad (1)$$

$$w_{random} = \frac{1}{V + \mathcal{I}^2} \quad (2)$$

$$w_{jrelative} [\%] = \frac{w_j}{\sum_{j=1}^k w_j} \times 100 \quad (3)$$

$V$  = Varianz der eingeschlossenen Studie

bei  $j = 1, \dots, k$  eingeschlossenen Studien

$\mathcal{I}^2$  = Interstudienvarianz

### *Metaanalyse von Überlebenszeitstudien*

Um die Ergebnisse einzelner Überlebenszeitstudien auf einer linearen Skala zusammenzufassen, wird die Datensynthese anhand der jeweils studienspezifischen  $\log HR$ s bzw.  $\log(HR)$  und ihrer Varianzen  $var(\log(HR))$  durchgeführt.<sup>9</sup>

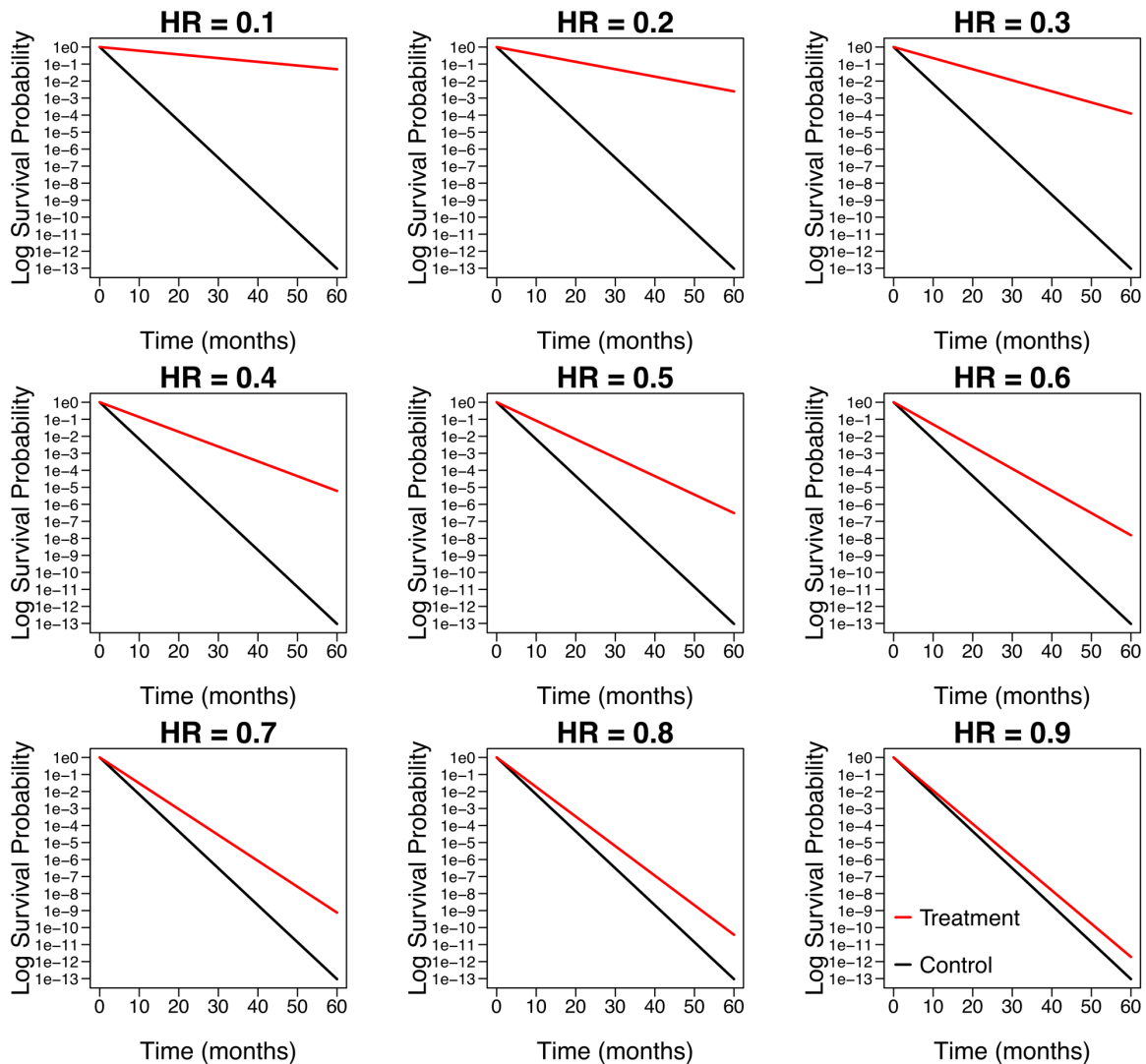


Abb. 3: Die Log-Transformation ermöglicht die lineare Darstellung von Hazard Ratios. Abstände zwischen benachbarten y-Achsenwerten entsprechen nunmehr Zehnerpotenzen.

Jedoch werden diese in der medizinischen Literatur oft nicht ausreichend berichtet.<sup>10–12</sup> Hierin begründen sich zahlreiche Bestrebungen  $\log$  HRs und ihre dazugehörige Varianzen aus aggregierten Daten publizierter Studien zu schätzen. Denn die Evidenzsynthese in Form systematischer Reviews und Metaanalysen stellt eine elementare Arbeitstechnik zur Bewertung von Wirksamkeit und Sicherheit medizinischer Maßnahmen und ist in diesem Rahmen von zentraler Bedeutung.<sup>13</sup> Dementsprechend wird eine große Anzahl von Metaanalysen jährlich publiziert, wobei die Zahl kontinuierlich steigt und zuletzt fast 30000 im Jahre 2019 betrug.<sup>14</sup> Dabei konnte nachgewiesen werden, dass sich über 95% der Metaanalysen auf aggregierte Daten stützen.<sup>15</sup> Hieraus lässt sich schlussfolgern, dass gerade im Rahmen der Überlebenszeitanalyse, wo relevante Kenngrößen in hohem Maße unberichtet bleiben, approximative Schätzungen zwangsläufig an der Tagesordnung sind.

### *Aktueller Stand der Forschung*

Parmar und Kollegen schlugen verschiedene Methoden zur indirekten Schätzung des *log HR* aus der Gesamtereignisanzahl einer Studie und der Log-Rang-Teststatistik mit dazugehörigem P-Wert, sowie der Varianz des *log HR* aus ggf. berichteten symmetrischen 95% Konfidenzintervallen vor.<sup>9</sup> Dabei wiesen sie bereits auf eine Unzuverlässigkeit der Schätzung im Falle unbalanzierter Designs hin. Außerdem schlugen die Autoren einen Ansatz zur Rekonstruktion des *log HR* aus Kaplan-Meier Kurven vor, der von Tudor und Williamson weiterentwickelt wurde.<sup>16,17</sup> Da die vorgenannten Daten – Ereigniszahlen, Log-Rang-Teststatistik, P-Wert – jedoch ebenfalls nur selten ausreichend berichtet werden, schlugen Michiels und Kollegen die Schätzung anhand medianer Überlebenszeiten vor, was sich allerdings als sehr unpräzise erwies.<sup>18</sup>

Tierney et al. fassten schließlich alle bis dato publizierten Vorschläge zusammen, definierten 11 Szenarien individueller Datenverfügbarkeit und entwickelten ein entsprechendes fallbasiertes Berechnungstool.<sup>19</sup> Die Arbeit der Autoren umfasst zwei direkte Berechnungswege und sieben indirekte Schätzmethoden der *log HRs* und ihrer Varianzen aus aggregierten Daten publizierter Studien. Dabei sind die Autoren zum einen bemüht der Leserschaft möglichst benutzerfreundliche Formeln an die Hand zu geben, müssen aber zum anderen ebenso auf wichtige Limitationen hinweisen – insbesondere im Falle unbalanzierter Gruppengrößen, inkonstanter Zensierungsraten, stark variierender Studienpopulationen oder Eventraten. Eine systematische Untersuchung der Präzision der Schätzmethoden im Zuge unterschiedlicher Designparameter und entsprechend dafür ableitbare Empfehlungen fehlen jedoch.

Ein im Anschluss ebenfalls von Tierney durchgeführter Vergleich von aus 18 Metaanalysen berichteter *log HRs*, welche zum Zeitpunkt der Veröffentlichung jeweils auf Grundlage individueller Patientendaten berechnet worden waren, mit aus der jeweils korrespondierenden Primärliteratur durch die Autoren geschätzten *log HRs*, zeigte zunehmende Diskrepanzen bei sinkenden Fallzahlen, kleineren Effektmaßen und niedrigeren Eventraten.<sup>20</sup> Allerdings konnten die Autoren zum einen nur knapp die Hälfte der Studien einschließen und differenzierten zum anderen nicht zwischen den verschiedenen direkten und indirekten Schätzmethoden. Darüber hinaus wurden nur 1:1 randomisierte Studien analysiert. Eine erste Empfehlung zur hierarchischen



Ordnung zur Verfügung stehender Verfahren formulierte Hebert 2022.<sup>21</sup> Allerdings verglichen die Autoren nur vier verschiedene Methoden und untersuchten insbesondere nicht den potentiellen Einfluss der vorgenannten Designparameter auf die Performance der Schätzmethoden.

### *Wissenschaftlich-biometrische Fragestellung der vorliegenden Arbeit*

Die nach wie vor offene wissenschaftliche und dementsprechend zu beantwortende biometrische Fragestellung der vorliegenden Arbeit ist es den potentiellen Einfluss unterschiedlicher Ausprägungen von Designparametern auf die Genauigkeit der in der Literatur vorgeschlagenen Schätzmethoden der *log HR* im Kontext unzureichend berichteter Überlebenszeitstudien systematisch zu quantifizieren. Entsprechend der wissenschaftlichen Fragestellung sollen zum einen Performancegrenzen verfügbarer Schätzmethoden der *log HR* im Kontext unzureichend berichteter Überlebenszeitstudien systematisch abgeleitet und zum anderen darauf basierende Anwendungsempfehlung bzw. Vorschläge zur situationsspezifischen Wahl der Schätzmethoden formuliert werden.

### *Biometrisch-biostatistische Bedeutung von Simulationsstudien*

Um die Präzision biometrischer Schätzmethoden systematisch in einem standardisierbaren Untersuchungsansatz zu analysieren, eignen sich Simulationsstudien. Diese sind computerbasierte Experimente, in deren Rahmen Datensätze aus Pseudozufallszahlen vorgegebener Verteilungen generiert werden, um die Performance biostatistischer Methoden in mannigfaltig variierbaren Szenarien im frequentistischen Sinne zu beurteilen.<sup>22</sup> Im Gegensatz zu realen biomedizinischen Daten, deren zugrundeliegende Struktur im Regelfall unbekannt ist und dem Wissenschaftler verborgen bleibt, ermöglicht die aktive Festlegung determinierender Faktoren, wie bspw. der Verteilungsparameter des Pseudozufallszahlengenerators, dem Experimentator das Verhalten biostatistischer Methoden im Kontext bekannter Szenarien zu untersuchen.<sup>23</sup> Mit dem Ziel die Durchführung, Auswertung und Kommunikation von Simulationsstudien zu standardisieren, führten Morris et al. 2019 die „ADEMP“-Struktur ein, die breite Akzeptanz erfuhr und folgende Bausteine umfasst:<sup>24</sup>

- **Aims** = Motivierende Fragestellung der Simulationsstudie:  
Die Identifikation neuer Auswertungsmethoden für bekannte Szenarien oder die Evaluation bereits bestehender Auswertungsmethoden in extremen Szenarien können Beispiele für Ausgangsfragestellungen darstellen.
- **Data-generating mechanisms (DGM)** = Eigenschaften der Datensatzerzeugung:  
Die der Simulationsstudie zugrundeliegenden Datensätze werden üblicherweise mithilfe eines Pseudozufallszahlengenerators erzeugt, wobei entweder wiederholte Ziehungen aus prädefinierten Verteilungen („parametric draws“), oder aber Stichprobenwiederholungen ohne Verteilungsannahme („resampling“) durchgeführt werden. Dabei können einzelne Parameter der Datensatzerzeugung faktoriell variiert werden, um ihren Einfluss zu untersuchen.
- **Estimands** = Gesamtheit der zu schätzenden Effekte / Ziele der Simulation:  
In Abhängigkeit von der Fragestellung der Simulationsstudie können z.B. Präzision und Richtigkeit von Schätzmethoden, Fehlerraten, Fallzahlkalkulationen, oder Designparameter von Interesse sein.
- **Methods** = Zu untersuchende Methoden:  
Typischerweise werden biostatistische Methoden entweder auf ihre Funktionalität („proof-of-concept“) oder auf ihre Limitationen („robustness“) hin untersucht bzw. verglichen.
- **Performance measures** = Messgrößen zur quantitativen Beurteilung der untersuchten Methoden:  
Maßzahlen, wie die mittlere Abweichung von Schätzmethoden im Vergleich zum wahren Wert („Bias“), werden a priori definiert, um die Ergebnisse der Simulationsstudie quantifizierbar zu machen. In Abhängigkeit von der Anzahl der Simulationsdurchläufe steigt dabei die Genauigkeit der zugrundeliegenden Berechnung („Monte-Carlo Fehler“).

## Material und Methoden

### *Szenarien der vorgelegten Simulationsstudie*

Eine Simulationsstudie nach der o.g. ADEMP-Struktur sollte durchgeführt werden, um den Grad der Übereinstimmung bzw. Abweichung jeweils geschätzter  $\log(HR)$  von den wahren  $\log(HR_{true})$  im Zweigruppenvergleich unter proportionalen und konstanten Ausfallraten in Abhängigkeit folgender faktoriell iterierter Parameter im Kontext exponentialverteilter Überlebenszeiten zu untersuchen:

- (1) Vorgegebene Hazard Ratios:  $HR_{true} = h_1(t)/h_0(t) = \{0.1, 0.3, 0.5, 0.7, 0.9\}$
- (2) Vorgegebene Gesamtfallzahlen:  $n_{obs} = \{100, 500, 1000\}$
- (3) Vorgegebene Allokationsraten:  $r = \{1:4, 1:2, 1:1, 2:1, 4:1\}$
- (4) Vorgegebene Baseline Hazards:  $\lambda = \{0.25, 0.5, 0.75\}$
- (5) Vorgegeben Zensierungsrate:
  - a. Keine zusätzliche Zensierung während des Beobachtungszeitraumes
  - b.  $C_{exp} = \text{Exp}(0.5)$
  - c.  $C_{unif} = \text{Unif}(0, \tau)$

mit Wahl von  $\tau$  zum Erreichen einer 50% Zensierungsrate

Im Folgenden werden die Schritte der im Rahmen der Konzipierung der vorliegenden Simulationsstudie konsentierten Aspekte gemäß der ADEMP-Struktur einzeln erläutert. Zudem sind die entsprechenden Teilaspekte in Tabelle 1 aufgelistet:

Tabelle 1: ADEMP-Struktur der vorliegenden Simulationsstudie

Aims:	Beurteilung der Performance approximativer Berechnungsmethoden für $\log(HR)$ und $\text{var}(\log(HR))$ unzureichend berichteter Überlebenszeitstudien.	Etablieren einer hierarchischen Wertung der Schätzmethoden und Formulieren szenarienspezifischer Verwendungsempfehlungen.
DGM:	Exponentialverteilte Überlebenszeitdaten im Zweigruppenvergleich:	$X_1, X_2, \dots, X_{n_x} \sim \text{Exp}(\lambda)$ mit $X_1$ bis $X_{n_x}$ Überlebenszeiten der Kontrollgruppe und $n_x$ Probanden in der Kontrollgruppe $Y_1, Y_2, \dots, Y_{n_y} \sim \text{Exp}(\lambda * HR)$ mit $Y_1$ bis $Y_{n_y}$ Überlebenszeiten der Experimentalgruppe und $n_y$ Probanden in der Experimentalgruppe
	Mit faktorieller Variation folgender Designparameter:	Hazard Ratio: $HR = \{0.1, 0.3, 0.5, 0.7, 0.9\}$

		<p><i>Gesamtfallzahl:</i>  <math>n_{obs} = n_x + n_y = \{100, 500, 1000\}</math></p> <p><i>Allokationsrate</i>  <math>r = n_x : n_y \{1:4, 1:2, 1:1, 2:1, 4:1\}</math></p> <p><i>Baseline Hazard:</i>  <math>\lambda = \{0.25, 0.5, 0.75\}</math></p> <p><i>Zusätzliche Zensierungen:</i>  <i>Keine zusätzlichen Zensierungen</i>  <math>C_{exp} \sim \text{Exp}(0.5)</math>  <math>C_{unif} \sim U(0, \tau)</math>                      mit <math>\tau = 2 * X</math> bzw. <math>Y</math>                      zum Erreichen einer ca. 50%-igen Zensierungsrate</p>
	Mit 1000 Wiederholungen pro Szenario ausgehend von einem Startwert von 30121988 für die Pseudozufallszahlen:	<p><math>n_{sim} = 1000</math>  <math>seed = 30121988</math></p>
Estimands:	Genauigkeit der approximativen Berechnungsmethoden für $\log(HR)$ , bestehend aus Abweichung und Streuung.	Resultierende Gewichtung der Methoden auf Grundlage der Varianzschätzer im Kontext der inversen Varianz-Methode.
Methods:	Cox-Regression	Methoden 1 bis 9 aus Tierney et al.
Performance Measures:	Bias und MCSE, Varianz, MSE, RMSE	Inverse Varianzen und prozentuale Gewichte

***Aims: motivierende Fragestellung der vorliegenden Simulationsstudie***

Die im Methodenpapier von Tierney et al. zusammengetragenen direkten und indirekten Berechnungsmethoden für  $\log(HR)$  und  $var(\log(HR))$  unzureichend berichteter Überlebenszeitstudien sollen in Abhängigkeit systematisch variiertes potentiell einflussreicher Designparameter an mittels Pseudozufallszahlen generierter simulierter *time-to-event* Datensätze evaluiert werden. Ziel dieser Simulationsstudie ist es i) mithilfe der so gewonnenen Erkenntnisse im frequentistischen Sinne Schlussfolgerungen bzgl. der Erwartungstreue und Präzision der genannten Methoden zu inferieren, um ii) schlussendlich eine datenbasierte hierarchische Wertung der Genauigkeit der selbigen zu formulieren.

### *Data-generating mechanisms: Daten-generierende Modelle in der vorliegenden Simulationsstudie*

Zunächst wurde eine Parametermatrix durch faktorielle Kombination der zu variierenden Designparameter mit jeweils 1000 Wiederholungen pro Szenario angelegt und der Startwert 30121988 für die Pseudozufallszahlen festgelegt:

```
nsim = 1000 # number of repetitions
set.seed(30121988) # random seed key

grid <- expand.grid(sim = 1:nsim, # 1000 repetitions
  hr = c(0.1,0.3,0.5,0.7,0.9), # true hazard ratios
  nobs = c(100,500,1000), # trial sizes
  r_c = round_half_up(c(4,2,1,1/2,1/4),2), # allocation ratios
  r_t = 1,
  lambda_c = c(0.25,0.5,0.75), # baseline event rates
  censoring = c("none","exp0.5","unif0.5")) # censoring rates
```

Im Anschluss wurde eine Datenmatrix, deren Zeilen jeweils einer simulierten Beobachtung, also einem Patienten, entsprachen, bestehend aus vier Kovariablen, mithilfe der folgenden Pseudozufallszahlengeneratoren erzeugt:

#### **1) Binäre Kovariable "Treatment"**

```
r = r_c + r_t # corresponds to allocation ratios
group_0 = (nobs/r)*r_c # unrounded size of control arm
group_1 = (nobs/r)*r_t # unrounded size of treatment arm
group_c = round_half_up(group_0,0) # rounded size of control arm
group_t = nobs-group_c # rounded size of treatment arm
treatment = c(rep(0,group_c), rep(1,group_t)) # treatment indicator
```

#### **2) Univariat exponentialverteilte Kovariable "Survivaltime"**

```
lambda_t = lambda_c*HR # event rate of treatment arm
survtimec = rexp(n=group_c, rate=lambda_c) # survival time of control arm
survtimet = rexp(n=group_t, rate=lambda_t) # survival time of treatment arm
survivaltime = c(survtimec, survtimet) # survival times of both arms
```

### 3) Univariat exponential-/uniformverteilte Kovariable "Censoring" und

### 4) Univariat exponentialverteilte Kovariable "Observed" in Abhängigkeit von der Zensierungsrate

```
duration = 60 # simulated trial observation time in months
accrualt = runif(n=nobs,min=0,max=12) # uniformly distributed accrual time
followupt = duration-accrualt # follow-up time

  if (censoring == "none") { # no additional censoring

    censoring = ifelse(survivaltime > followupt,0,1) # censoring at trial end only
observed = pmin(survivaltime, followupt) # observed time (censored/event)
    } else if (censoring == "exp0.5") { # exponential distribution
    C <- rexp(n=nobs, rate=0.5) # exponentially distributed variable
    censoring <- as.numeric(survivaltime <= C) # censoring status
observed = pmin(C, survivaltime, followupt) # observed time (censored/event)
    } else if (censoring == "unif0.5") {# uniform distribution
    C <- runif(n=nobs,min=0,max=2*survivaltime) # uniformly distributed variable
    censoring <- as.numeric(survivaltime <= C) # censoring status
observed = pmin(C, survivaltime, followupt) # observed time (censored/event)
  }
```

Der Matrixstruktur der Daten-generierenden Modelle entsprechend erfolgte die Simulation zeilenweise, wobei eine Zeile der Datenmatrix jeweils einer simulierten Überlebenszeitstudie entsprach und jeweils 1000 Wiederholungen bzw. Zeilen einem Szenario (also einer faktoriellen Kombination der zu variierenden Designparameter).

*Estimands: zu schätzende Zielgrößen der vorliegenden*

*Simulationsstudie*

Um die Genauigkeit der direkten und indirekten Berechnungsmethoden für  $\log(HR)$  zu evaluieren sollten i) die mittlere Abweichung der  $\log(HR)$ -Schätzer von den wahren  $\log(HR_{true})$ -Werten (*Bias*) und ihre Präzision in Form der Streuung (*Precision*) quantifiziert werden. Um die  $var(\log(HR))$  und etwaige Gewichtung zu beurteilen sollten ii) die inversen der  $var(\log(HR))$ -Schätzer evaluiert werden.

Im Nachgang sollte auf Grundlage der vorliegenden Genauigkeiten der Schätzmethoden eine hierarchische Ordnung der selbigen im Sinne einer szenarienspezifischen Verwendungsempfehlung formuliert werden.

### ***Methods: beobachtete log HRs und Varianzen***

Als Bestandteil der faktoriellen Designparameter der vorliegenden Simulationsstudie wurden die vorgegebenen Hazard Ratios von 0.1 bis 0.9 in Schritten von 0.2 iteriert:

$$HR_{true} = h_1(t)/h_0(t) = \{0.1, 0.3, 0.5, 0.7, 0.9\}$$

Da sich die überwiegende Mehrheit publizierter Überlebenszeitanalysen der semi-parametrischen *Cox-Regression* bedient,<sup>12</sup> wurde basierend auf den vier Kovariablen der per Pseudozufallszahlengeneratoren erzeugten Datenmatrix (*Treatment*, *Survivaltime*, *Censoring*, *Observed*, s.o.) zunächst eine solche mithilfe der „coxph“-Funktion aus dem R-Paket „survival“ berechnet (Version 3.2-3; Therneau, 2020), um dann die in etwaigen fiktiven Publikationen beobachteten  $\log(HR_{obs})$  und  $\text{var}(\log(HR_{obs}))$ , sowie 95% Konfidenzintervalle und P-Werte dementsprechend zu bestimmen:<sup>25</sup>

### **Cox-Regression**

```
coxph <- coxph(Surv(observed, censoring)~treatment, data) # Cox PH model

# Corresponding hazard ratio rounded to two decimal places:
HRobs = round_half_up(exp(coef(coxph))[[1]],2)

# Lower 95% confidence interval limit rounded to two decimal places:
L95 = round_half_up(exp(confint(coxph))[[1]],2)

# And limited to 0.01:
Low95obs <- ifelse(L95 == 0, 0.01, L95)

# Upper 95% confidence interval limit rounded to two decimal places:
Up95obs = round_half_up(exp(confint(coxph))[[2]],2)

# Corresponding p-value rounded to three decimal places (p < 0.001):
pvals = round_half_up(summary(coxph)$sctest['pvalue'][[1]],3)

# And limited to 0.0001:
pvalobs <- ifelse(pvals == 0, 0.0001, pvals)
```

```
# Corresponding log(HRobs) :
lnHRobs = coef(summary(coxph))[, 'coef']

# Corresponding var(log(HRobs)) :
lnVobs = diag(coxph$var)
```

Dabei wurden Hazard Ratios und Konfidenzintervallgrenzen entsprechend des *American Psychological Association (APA)* „Numbers and Statistics Guide“ auf zwei Nachkommastellen und P-Werte auf drei Nachkommastellen gerundet, um den üblichen Konventionen und damit dem Reporting etwaiger fiktiver Publikationen zu entsprechen.<sup>26</sup>

### Methoden 1 bis 9 nach Tierney

Im Anschluss wurden die direkten und indirekten Berechnungsmethoden für  $\log(HR)$  und  $\text{var}(\log(HR))$  aus Tierney et al. auf denselben Datensatz angewendet.<sup>19</sup> Diese umfassen fünf Methoden zur approximativen Berechnung des Hazard Ratio und sechs Methoden zur approximativen Berechnung der Log-Rang-Varianz.

#### Berechnungen der HR nach Tierney:

$$HR = \frac{\text{observed events research} / \text{logrank expected events research}}{\text{observed events control} / \text{logrank expected events control}} \quad (4)$$

$$HR = \exp \left[ \frac{\text{observed} - \text{expected events research} (O - E)}{\text{variance of the logrank test statistic}} \right] \quad (5)$$

where O-E is the logrank observed minus expected events

$$O - E = \sqrt{\frac{\text{observed events research} \times \text{observed events control}}{\text{total observed events}}} \times \Phi^{-1} \left( 1 - \frac{p}{2} \right) \quad (6)$$

where p is the reported p-value of the study to be included  
then using equation (5)

$$O - E = \frac{1}{2} \times \sqrt{\text{total observed events}} \times \Phi^{-1} \left( 1 - \frac{p}{2} \right) \quad (7)$$

where p is the reported p-value of the study to be included  
then using equation (5)

$$O - E = \sqrt{\frac{\text{total observed events} \times \text{analyzed research} \times \text{analyzed control}}{\text{analyzed research} + \text{analyzed control}}} \times \Phi^{-1} \left( 1 - \frac{p}{2} \right) \quad (8)$$

where p is the reported p-value of the study to be included  
then using equation (5)



*Berechnungen der log var nach Tierney:*

$$\text{logrank variance} = \frac{1}{(1/\text{expected events research}) + (1/\text{expected events control})} \quad (9)$$

$$\text{logrank variance} = \frac{\text{observed} - \text{expected evens research } (O - E)}{\log(HR)} \quad (10)$$

$$\text{logrank variance} = 1 / \left[ \frac{\log(\text{upper 95\% CI limit}) - \log(\text{lower 95\% CI limit})}{2 \times \Phi^{-1}(1 - \alpha/2)} \right]^2 \quad (11)$$

where  $\alpha$  is the significance level stated in the study to be included

$$\text{logrank variance} = \frac{\text{observed events research} \times \text{observed events control}}{\text{total observed events}} \quad (12)$$

$$\text{logrank variance} = \frac{\text{total observed events}}{4} \quad (13)$$

$$\text{logrank variance} = \frac{\text{total observed events} \times \text{analyzed research} \times \text{analyzed control}}{(\text{analyzed research} + \text{analyzed control})^2} \quad (14)$$

Dem Kontext der Metaanalyse Rechnung tragend wurden die o.g. Formeln zur approximativen Bestimmung der Hazard Ratio und Log-Rang-Varianz um die nachfolgenden Umwandlungen in die korrespondierenden  $\log(HR)$  und  $\text{var}(\log(HR))$  direkt erweitert:

*Berechnung der log(HR) und var(log(HR)):*

$$\log(HR) \hat{=} \log_e(HR) \quad (15)$$

$$\text{var}(\log(HR)) = 1 / \log \text{var} \quad (16)$$

Die aufgeführten direkten und indirekten Berechnungsmethoden für  $\log(HR)$  und  $\text{var}(\log(HR))$  setzen sich wie folgt zu neun Methoden, abhängig vom Informationsgehalt der berichteten Studien, zusammen:

- **Method 1** (Anzahl beobachteter und erwarteter Ereignisse oder Ausfallraten von Kontroll- und Experimentalgruppe werden berichtet):

$$\ln HR_{dir1}: ( 4 )$$

$$\text{var}(\ln HR_{dir1}): ( 9 )$$

- **Method 2** (Differenz zwischen beobachteten und unter Nullhypothese erwarteten Ereignissen und Varianz der Log-Rang-Teststatistik werden berichtet):

$$\ln HR_{dir2}: ( 5 )$$

$$\text{var}(\ln HR_{dir2}): ( 10 )$$

- **Method 3** (Hazard Ratio und 95% Konfidenzintervall werden berichtet):

$$\ln HR_{indir1}: HR_{Obs}$$

$$\text{var}(\ln HR_{indir1}): ( 11 )$$

- **Method 4** (Hazard Ratio und Anzahl der Ereignisse von Kontroll- und Experimentalgruppe bei einer Allokationsrate von 1 zu 1 werden berichtet):

$$\ln HR_{indir2}: HR_{Obs}$$

$$\text{var}(\ln HR_{indir2}): ( 12 )$$

- **Method 5** (Hazard Ratio und Gesamtanzahl der Ereignisse bei einer Allokationsrate von 1 zu 1 werden berichtet):

$$\ln HR_{indir3}: HR_{Obs}$$

$$\text{var}(\ln HR_{indir3}): ( 13 )$$

- **Method 6** (Hazard Ratio und Gesamtanzahl der Ereignisse werden berichtet, wobei nicht alle randomisierten Patienten ausgewertet wurden):

$$\ln HR_{indir4}: HR_{Obs}$$

$$\text{var}(\ln HR_{indir4}): ( 14 )$$

- **Method 7** (*P-Wert und Anzahl der Ereignisse von Kontroll- und Experimentalgruppe bei einer Allokationsrate von 1 zu 1 werden berichtet*):

$$\ln HR_{\text{indir}5}: ( 6 )$$
$$\text{var}(\ln HR_{\text{indir}5}): ( 12 )$$

- **Method 8** (*P-Wert und Gesamtanzahl der Ereignisse bei einer Allokationsrate von 1 zu 1 werden berichtet*):

$$\ln HR_{\text{indir}6}: ( 7 )$$
$$\text{var}(\ln HR_{\text{indir}6}): ( 13 )$$

- **Method 9** (*P-Wert und Gesamtanzahl der Ereignisse bei einer Allokationsrate ungleich 1 zu 1 werden berichtet*):

$$\ln HR_{\text{indir}7}: ( 8 )$$
$$\text{var}(\ln HR_{\text{indir}7}): ( 14 )$$

### Berechnungsmethoden nach Tierney umgesetzt in RStudio

Der Struktur der Datenmatrix entsprechend erfolgte die Ausweitung zeilenweise, mit einer simulierten Überlebenszeitstudie pro Zeile. Die nachfolgenden Ergebnisvektoren wurden daher zeilenweise pro Simulationsdurchgang befüllt, wobei jede  $\log(HR)$ - und  $\text{var}(\log(HR))$ -Schätzmethode auf dieselbe simulierte Überlebenszeitstudie angewendet wurde.

```
# Total number of simulated patients:

Obs = nrow(data)

# Number of simulated patients per study arm:

Res = sum(data$treatment)

Con = Obs-Res

# Logrank statistic:

logrank <- survdiff(Surv(observed, censoring)~treatment, data)

m <- survfit(Surv(observed, censoring)~treatment, data)
```

– Material und Methoden –

```
# METHOD 1 #

lnHRdir1 = log((logrank$obs[2]/logrank$exp[2])/(logrank$obs[1]/logrank$exp[1]))

lnVdir1 = 1/(1/(1/logrank$exp[2] + 1/logrank$exp[1]))

# METHOD 2 #

O_Edir = logrank$obs[2]-logrank$exp[2]

Vdir2 = O_Edir^2/logrank$chisq

lnHRdir2 = log(exp(O_Edir/Vdir2))

lnVdir2 = 1/Vdir2

# METHOD 3 #

lnHRindir1 = lnHRobs

lnVindir1 = ((log(Up95obs)-log(Low95obs))/(2*qnorm(1-0.05/2)))^2

# METHOD 4 #

lnHRindir2 = lnHRobs

lnVindir2 = 1/((logrank$obs[2]*logrank$obs[1])/(logrank$obs[2]+logrank$obs[1]))

# METHOD 5 #

lnHRindir3 = lnHRobs

lnVindir3 = 1/((logrank$obs[2]+logrank$obs[1])/4)

# METHOD 6 #

lnHRindir4 = lnHRobs

lnVindir4 = 1/(((logrank$obs[2]+logrank$obs[1])*Con*Res)/(Con+Res)^2)

# METHOD 7 #

O_Eindir5 = -(sqrt((logrank$obs[2]*logrank$obs[1]/

(logrank$obs[2]+logrank$obs[1]))))* (qnorm(1-pvalobs/2))
```

```
Vindir5 = logrank$obs[2]*logrank$obs[1]/(logrank$obs[2]+logrank$obs[1])
lnHRindir5 = log(exp(O_Eindir5/Vindir5))
lnVindir5 = 1/Vindir5

# METHOD 8 #

O_Eindir6 = -0.5*sqrt((logrank$obs[2]+logrank$obs[1]))*(qnorm(1-pvalobs/2))
Vindir6 = (logrank$obs[2]+logrank$obs[1])/4
lnHRindir6 = log(exp(O_Eindir6/Vindir6))
lnVindir6 = 1/Vindir6

# METHOD 9 #

O_Eindir7 = -sqrt((logrank$obs[2]+logrank$obs[1])*Con*Res)/
              (Con+Res)*(qnorm(1-pvalobs/2))
Vindir7 = ((logrank$obs[2]+logrank$obs[1])*Con*Res)/(Con+Res)^2
lnVindir7 = 1/Vindir7
lnHRindir7 = log(exp(O_Eindir7/Vindir7))
```

Dabei muss auf Unterschiede in der Berechnung der Hazard Ratios auf Grundlage der Log-Rang-Teststatistik versus der Cox-Regression hingewiesen werden, die im Nachgang diskutiert werden.<sup>25,27,28</sup>

### *Performance measures 1: Erwartungstreue und Verzerrtheit der Punktschätzer für $\log(HR)$ in der vorliegenden Simulationsstudie*

#### **Verzerrtheit (Bias)**

Neben der Ausgabe der arithmetischen Mittelwerte der Punktschätzer für  $\log(HR)$  sollen Erwartungstreue versus systematischer Fehler der approximativen Schätzmethoden jeweils als mittlere Abweichung vom „wahren“ vorgegebenen  $HR_{true}$  bzw.  $\log(HR_{true})$  als Maß für die Verzerrtheit quantifiziert werden (*Bias*):<sup>24</sup>

Messgröße	Definition	Schätzer	Monte Carlo Standardfehler
Bias	$E[\hat{\theta}] - \theta$	$\frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} \hat{\theta}_i - \theta$	$\sqrt{\frac{1}{n_{sim} \times (n_{sim} - 1)} \sum_{i=1}^{n_{sim}} (\hat{\theta}_i - \bar{\theta})^2}$

$n_{sim}$  = Anzahl der Simulationsdurchläufe  
 $\hat{\theta}_i$  = geschätztes  $\log(HR)$  pro Simulationsdurchlauf

### Monte Carlo Standardfehler des Bias (MCSE)

Der dazugehörige Monte Carlo Standardfehler (MCSE) misst dabei die der stochastischen Natur der Simulationsstudie geschuldete Unsicherheit, die durch die endliche Zahl der Wiederholungen begründet ist. Er ist somit ein Maß für die inhärente Streuung der Simulationsergebnisse selbst und somit für die Präzision der Simulation.

### Varianz der Punktschätzer

Die Varianz der Punktschätzer gibt Auskunft über die Streuung der Schätzmethoden:

Messgröße	Definition	Schätzer
Varianz	$\text{Var}(\hat{\theta})$	$\frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} (\hat{\theta}_i - \bar{\theta})^2$

$n_{sim}$  = Anzahl der Simulationsdurchläufe  
 $\hat{\theta}_i$  = geschätztes  $\log(HR)$  pro Simulationsdurchlauf

### Der mittlere quadrierte Fehler (MSE)

Der mittlere quadrierte Fehler (MSE) entspricht der erwarteten quadratischen Abweichung und setzte sich aus Varianz und Bias zusammen, ist somit ein kombiniertes Maß für Verzerrtheit und Streuung (Genauigkeit):

Messgröße	Definition	Schätzer
MSE	$E[(\hat{\theta} - \theta)^2]$	$\frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} (\hat{\theta}_i - \theta)^2$

$n_{sim}$  = Anzahl der Simulationsdurchläufe  
 $\hat{\theta}_i$  = geschätztes  $\log(HR)$  pro Simulationsdurchlauf

### Die Wurzel des mittleren quadrierten Fehlers (RMSE)

Um die Interpretation des mittleren quadrierten Fehlers durch Wiederherstellung des ursprünglichen Skalenniveaus der Schätzer zu erleichtern, wird die Wurzel gebildet (RMSE):

Messgröße	Definition	Schätzer
RMSE	$\sqrt{MSE}$	$\sqrt{\frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} (\hat{\theta}_i - \theta)^2}$

$n_{sim}$  = Anzahl der Simulationsdurchläufe  
 $\hat{\theta}_i$  = geschätztes  $\log(HR)$  pro Simulationsdurchlauf

### calculate\_performances()

Um alle vorgenannten Berechnungen zu ermöglichen, wurde die „calc\_absolute“-Funktion aus dem „simhelpers“ Paket von Joshi und Pustejovsky umprogrammiert:<sup>29</sup>

```
calculate_performances <- function(res_dat, estimates, true_param,
perfm_criteria = c("Estimate", "Variance", "Bias", "MCSE", "MSE", "RMSE"))
{
  estimates <- res_dat %>% # define data set
  dplyr::filter(!is.na({{estimates}})) %>% # exclude NAs
  dplyr::pull({{estimates}}) # pull estimates from data set
  Iterations <- length(estimates) # number of iterations (nsim)
  true_param <- res_dat %>% dplyr::pull({{true_param}}) # pull true parameters
  true_param <- true_param[1] # set true parameters
  # calculate sample stats:
  t_bar <- mean(estimates) # mean of estimates
  var_t <- var(estimates) # variance of estimates
  bias <- t_bar - true_param # bias of estimates
  s_t <- sd(estimates) # standard deviation of estimates
  mse <- mean((estimates - true_param)^2) # calculate mse
  # initialize tibble:
  dat <- tibble::tibble(Iterations = Iterations)
```

```
if("Estimate" %in% perfm_criteria){ #POINT ESTIMATES

  dat <- dat %>%

  dplyr::mutate(Theta = t_bar)

}

if("Variance" %in% perfm_criteria){ #VARIANCE OF POINT ESTIMATES

  dat <- dat %>%

  dplyr::mutate(Variance = var_t)

}

if("Bias" %in% perfm_criteria){ #BIAS OF POINT ESTIMATES

  dat <- dat %>%

  dplyr::mutate(Bias = bias)

}

if("MCSE" %in% perfm_criteria){ #MCSE OF BIAS

  dat <- dat %>%

  dplyr::mutate(MCSE = sqrt(var_t / Iterations),

                Lower = bias-qnorm(0.975)*MCSE,

                Upper = bias+qnorm(0.975)*MCSE)

}

if("MSE" %in% perfm_criteria){ #MSE

  dat <- dat %>%

  dplyr::mutate(MSE = mse)

}

if("RMSE" %in% perfm_criteria){ #RMSE

  dat <- dat %>%

  dplyr::mutate(RMSE = sqrt(mse))

}

return(dat)

}
```



***Performance measures 2: Punktschätzer für  $\text{var}(\log(HR))$  und jeweiliger prozentualer Gewichtanteil in der vorliegenden Simulationsstudie***

Neben der Beurteilung der Genauigkeit der  $\log(HR)$ -Punktschätzer (Formeln 4 bis 8), sollen die  $\text{var}(\log(HR))$ -Punktschätzer (Formeln 9 bis 14) inkl. der dazugehörigen simulationsbedingten Standardfehler ausgegeben werden (MCSE).

**calculate\_variances()**

Um die Varianzschätzer und Standardfehler auszugeben, wurde die „calc\_absolute“-Funktion aus dem „simhelpers“ Pakets nochmals wie folgt umprogrammiert:<sup>29</sup>

```
calculate_variances <- function(res_dat, estimates,
output = c("Estimate", "Variance", "MCSE"))
{
  estimates <- res_dat %>% # define data set
  dplyr::filter(!is.na({{estimates}})) %>% # exclude NAs
  dplyr::pull({{estimates}}) # pull estimates from data frame
  Iterations <- length(estimates) # number of iterations (nsim)
  # calculate sample stats:
  t_bar <- mean(estimates) # mean of estimates
  var_t <- var(estimates) # variance of estimates
  s_t <- sd(estimates) # standard deviation of estimates
  # initialize tibble:
  dat <- tibble::tibble(Iterations = Iterations)
  if("Estimate" %in% output){ #POINT ESTIMATES
    dat <- dat %>%
    dplyr::mutate(Theta = t_bar)
  }
  if("Variance" %in% output){ #VARIANCE OF POINT ESTIMATES
    dat <- dat %>%
    dplyr::mutate(Variance = var_t)
  }
}
```

```
if("MCSE" %in% output){ #STANDARD ERROR OF POINT ESTIMATES

  dat <- dat %>%

    dplyr::mutate(MCSE = sqrt(var_t / Iterations))

}

return(dat)

}
```

### Relative Studiengewichte

Der inversen Varianz-Methode im Festeffekte-Modell entsprechend, sollen auf Grundlage der beobachteten Varianzen der Cox-Regression und der approximativ berechneten  $var(\log(HR))$ -Punktschätzer (Formeln 9 bis 14) relative Studiengewichte für jede Schätzmethode nach den Formeln 1 und 3 bestimmt werden.

```
# Sum of weights:

Weights_all$total <- 1/Weights_all$lnVobs +
                    1/Weights_all$lnVdir1 +
                    1/Weights_all$lnVdir2 +
                    1/Weights_all$lnVindir1 +
                    1/Weights_all$lnVindir2 +
                    1/Weights_all$lnVindir3 +
                    1/Weights_all$lnVindir4

# Relative study weight for COX Model [in percent]:

Weights_all$COX = (1/Weights_all$lnVobs)/Weights_all$total*100

# Relative study weight for Method 1 [in percent]:

Weights_all$Dir1 = (1/Weights_all$lnVdir1)/Weights_all$total*100

# Relative study weight for Method 2 [in percent]:

Weights_all$Dir2 = (1/Weights_all$lnVdir2)/Weights_all$total*100

# Relative study weight for Method 3 [in percent]:

Weights_all$Indir1 = (1/Weights_all$lnVindir1)/Weights_all$total*100

# Relative study weight for Method 4 [in percent]:

Weights_all$Indir2 = (1/Weights_all$lnVindir2)/Weights_all$total*100
```

```
# Relative study weight for Method 5 [in percent]:  
Weights_all$Indir3 = (1/Weights_all$lnVindir3)/Weights_all$total*100  
  
# Relative study weight for Method 6 [in percent]:  
Weights_all$Indir4 = (1/Weights_all$lnVindir4)/Weights_all$total*100  
  
# Note:  
  
Methods 7 through 9 comprise the same variance estimators as Methods 4 through 6
```

So sollen die zeilenweise pro simuliertem Überlebenszeitdatensatz berechneten relativen Gewichte der Schätzmethoden und der Cox-Regression untereinander verglichen werden, wobei sich die inverse Gesamtvarianz aus den inversen Varianzen der Cox-Regression und der Schätzmethoden zusammensetzt. Neben der Beurteilung der Genauigkeit der Schätzmethoden, soll so die ihnen im Rahmen einer etwaigen Metaanalyse potentiell zuteilwerdende Gewichtung simuliert und beurteilt werden.

## Ergebnisse

### Erzeugte Ergebnisdatenmatrix

Nach 12.1 Tagen Rechenzeit enthielt die ausgegebene Ergebnisdatenmatrix 675000 Zeilen, entsprechend 675000 simulierten Überlebenszeitstudien mit 675 faktoriell variierten Szenarien, und 35 Spalten. In insgesamt 1356 Fällen hatte der Pseudozufallszahlengenerator ein beobachtetes Hazard Ratio von 0 erzeugt, in zwei Fällen ein beobachtetes Hazard Ratio von über 10. Nach Entfernung der vorgenannten Iterationen, verblieb ein Datensatz von 673642 simulierten Studien:

### Deskriptive Statistik der Ergebnisdatenmatrix “GRID”

Im Folgenden wird die Ausgabe der „summary“-Funktion, angewendet auf die Ergebnisdatenmatrix “GRID”, abgedruckt, welche eine spaltenweise Deskription der Ergebnisvektoren mit Minima und Maxima sowie 25%-Quantilen ausgibt:

```
# Simulation numbers, simulated trial sizes, allocation ratios, baseline hazards
##          sim              nobs              r_c              r_t              lambda_c
## Min.    : 1.0    Min.    : 100.0    Min.    :0.250    Min.    :1    Min.    :0.2500
## 1st Qu.: 250.0   1st Qu.: 100.0   1st Qu.:0.500   1st Qu.:1    1st Qu.:0.2500
## Median : 500.0   Median : 500.0   Median :1.000   Median :1    Median :0.5000
## 3rd Qu.: 750.0   3rd Qu.:1000.0   3rd Qu.:2.000   3rd Qu.:1    3rd Qu.:0.7500
## Max.    :1000.0   Max.    :1000.0   Max.    :4.000   Max.    :1    Max.    :0.7500

# Given true HR, censoring distributions, given log(true HR), number of scenario
##          hr              censoring              lnhr              Scenario
## Min.    :0.1000   none    :224999   Min.    : -2.3030   Min.    : 1.0
## 1st Qu.:0.3000   exp0.5  :223750   1st Qu.: -1.2040   1st Qu.:169.0
## Median :0.5000   unif0.5:224893   Median : -0.6930   Median :339.0
## 3rd Qu.:0.7000                                     3rd Qu.: -0.3570   3rd Qu.:507.0
## Max.    :0.9000                                     Max.    : -0.1050   Max.    :675.0

# Number of patients on control/research arms, observed HR and lower 95%CI bound
##          Con              Res              HRobs              Low95obs
## Min.    : 20.0    Min.    : 20.0    Min.    :0.0100    Min.    :0.0100
## 1st Qu.: 67.0    1st Qu.: 67.0    1st Qu.:0.2500    1st Qu.:0.1500
## Median :200.0    Median :200.0    Median :0.4800    Median :0.3300
```

- Ergebnisse -

```

## 3rd Qu.:400.0 3rd Qu.:400.0 3rd Qu.:0.7400 3rd Qu.:0.5400
## Max. :800.0 Max. :800.0 Max. :7.1100 Max. :2.1400
# Upper 95%CI bound, p-value of Cox Regression, log (observed HR) and variance
## Up95obs pvalobs lnHRobs lnVobs
## Min. : 0.0400 Min. :0.0001 Min. :-5.2619 Min. :0.004006
## 1st Qu.: 0.3600 1st Qu.:0.0001 1st Qu.:-1.3800 1st Qu.:0.012235
## Median : 0.6800 Median :0.0001 Median :-0.7297 Median :0.024175
## 3rd Qu.: 1.0300 3rd Qu.:0.0670 3rd Qu.:-0.3078 3rd Qu.:0.080226
## Max. :52.6400 Max. :1.0000 Max. : 1.9616 Max. :2.214175
# Results of Methods 1 and 2:
## lnHRdir1 lnVdir1 lnHRdir2 lnVdir2
## Min. :-4.5327 Min. :0.00400 Min. :-13.6626 Min. :0.004007
## 1st Qu.:-1.1874 1st Qu.:0.01026 1st Qu.: -1.5007 1st Qu.:0.011284
## Median :-0.6787 Median :0.02195 Median : -0.7455 Median :0.023314
## 3rd Qu.:-0.2999 3rd Qu.:0.07310 3rd Qu.: -0.3084 3rd Qu.:0.078886
## Max. : 1.9277 Max. :4.25325 Max. : 2.2348 Max. :4.434986
# Results of Methods 3 and 4:
## lnHRindir1 lnVindir1 lnHRindir2 lnVindir2
## Min. :-5.2619 Min. :0.003636 Min. :-5.2619 Min. :0.004000
## 1st Qu.:-1.3800 1st Qu.:0.012332 1st Qu.:-1.3800 1st Qu.:0.009352
## Median :-0.7297 Median :0.024358 Median :-0.7297 Median :0.019883
## 3rd Qu.:-0.3078 3rd Qu.:0.079352 3rd Qu.:-0.3078 3rd Qu.:0.062500
## Max. : 1.9616 Max. :1.986843 Max. : 1.9616 Max. :2.000000
# Results of Methods 5 and 6:
## lnHRindir3 lnVindir3 lnHRindir4 lnVindir4
## Min. :-5.2619 Min. :0.00400 Min. :-5.2619 Min. :0.004000
## 1st Qu.:-1.3800 1st Qu.:0.00800 1st Qu.:-1.3800 1st Qu.:0.009238
## Median :-0.7297 Median :0.01587 Median :-0.7297 Median :0.018731
## 3rd Qu.:-0.3078 3rd Qu.:0.04040 3rd Qu.:-0.3078 3rd Qu.:0.062500

```

– Ergebnisse –

```

## Max.      : 1.9616   Max.      :2.00000   Max.      : 1.9616   Max.      :3.125000

# Results of Methods 7 and 8:

##      lnHRindir5      lnVindir5      lnHRindir6      lnVindir6
## Min.      :-4.4925   Min.      :0.004000   Min.      :-3.8906   Min.      :0.00400
## 1st Qu.   :-0.6390   1st Qu.   :0.009352   1st Qu.   :-0.5246   1st Qu.   :0.00800
## Median    :-0.4149   Median    :0.019883   Median    :-0.3501   Median    :0.01587
## 3rd Qu.   :-0.2612   3rd Qu.   :0.062500   3rd Qu.   :-0.2461   3rd Qu.   :0.04040
## Max.      : 0.0000   Max.      :2.000000   Max.      : 0.0000   Max.      :2.00000

# Results of Method 9:

##      lnHRindir7      lnVindir7
## Min.      :-4.8632   Min.      :0.004000
## 1st Qu.   :-0.6146   1st Qu.   :0.009238
## Median    :-0.4067   Median    :0.018731
## 3rd Qu.   :-0.2612   3rd Qu.   :0.062500
## Max.      : 0.0000   Max.      :3.125000

```

*Performance measures der log(HR)-Schätzmethoden pro Hazard Ratio*

Tabelle 2 listet die Ergebnisse der Funktion *calculate\_performances()* angewendet auf die Ergebnisvektoren der Cox-Regression (Spalte „lnHRobs“) und der Methoden 1, 2, 7, 8 und 9 (Spalten „lnHRdir1“, „lnHRdir2“, „lnHRindir5“, „lnHRindir6“ und „lnHRindir7“), stratifiziert nach den vorgegebenen  $HR_{true} = \{0.1, 0.3, 0.5, 0.7, 0.9\}$ . Die Spalte „Methods“ bezeichnet die jeweils analysierte Auswertungsmethode und die Spalte „Iterations“ die Anzahl der zugrundeliegenden Simulationen. Die Spalte „Theta“ enthält die arithmetischen Mittelwerte der geschätzten  $\log(HR)$  und die Spalte „Variance“ die dazugehörigen empirischen Varianzen, wobei pro Methode und vorgegebenem Hazard Ratio über alle Szenarien mit jeweils  $n_{sim} = 1000$  pro Szenario gepoolt wurde (s. „Material und Methoden: Szenarien der vorgelegten Simulationsstudie“). Die Spalten „Bias“ und „MCSE“, sowie „Lower“ und „Upper“ enthalten die mittleren Abweichungen der Schätzmethoden vom vorgegebenen  $\log(HR_{true})$ , die dazugehörigen Standardfehler und ihre 95% Konfidenzintervalle. Die Spalten „MSE“ und „RMSE“ listen schließlich die mittleren quadrierten Fehler und ihre Wurzeln.

– Ergebnisse –

Method	Iterations	Theta	Variance	Bias	MCSE	Lower	Upper	MSE	RMSE
<i>Summary Statistics and Bias for Methods 1 through 9 with HR true = 0.1 [log(HR true) = -2.303]:</i>									
InHRobs	133747	-2.41090	0.12654	-0.10790	0.00097	-0.10981	-0.10600	0.13819	0.37174
InHRdir1	133747	-1.77653	0.15318	0.52647	0.00107	0.52437	0.52857	0.43035	0.65601
InHRdir2	133747	-2.76677	1.62501	-0.46377	0.00349	-0.47060	-0.45693	1.84008	1.35650
InHRindir5	133747	-0.85878	0.38353	1.44422	0.00169	1.44090	1.44754	2.46930	1.57140
InHRindir6	133747	-0.66268	0.16118	1.64032	0.00110	1.63817	1.64247	2.85182	1.68873
InHRindir7	133747	-0.74545	0.21434	1.55755	0.00127	1.55507	1.56003	2.64031	1.62490
<i>Summary Statistics and Bias for Methods 1 through 9 with HR true = 0.3 [log(HR true) = -1.204]:</i>									
InHRobs	134916	-1.26173	0.06748	-0.05773	0.00071	-0.05911	-0.05634	0.07081	0.26610
InHRdir1	134916	-1.08705	0.04585	0.11695	0.00058	0.11581	0.11810	0.05952	0.24398
InHRdir2	134916	-1.39466	0.23268	-0.19066	0.00131	-0.19323	-0.18809	0.26903	0.51868
InHRindir5	134916	-0.67691	0.13971	0.52709	0.00102	0.52510	0.52909	0.41754	0.64617
InHRindir6	134916	-0.58227	0.08994	0.62173	0.00082	0.62013	0.62333	0.47649	0.69028
InHRindir7	134916	-0.65358	0.11573	0.55042	0.00093	0.54861	0.55224	0.41869	0.64707
<i>Summary Statistics and Bias for Methods 1 through 9 with HR true = 0.5 [log(HR true) = -0.693]:</i>									
InHRobs	134982	-0.72711	0.05319	-0.03411	0.00063	-0.03534	-0.03288	0.05435	0.23313
InHRdir1	134982	-0.67303	0.03996	0.01997	0.00054	0.01891	0.02104	0.04036	0.20090
InHRdir2	134982	-0.76549	0.08387	-0.07249	0.00079	-0.07403	-0.07094	0.08912	0.29854
InHRindir5	134982	-0.53252	0.06074	0.16048	0.00067	0.15916	0.16179	0.08649	0.29410
InHRindir6	134982	-0.46892	0.04310	0.22408	0.00057	0.22297	0.22519	0.09331	0.30547
InHRindir7	134982	-0.52544	0.05452	0.16756	0.00064	0.16631	0.16880	0.08260	0.28740
<i>Summary Statistics and Bias for Methods 1 through 9 with HR true = 0.7 [log(HR true) = -0.357]:</i>									
InHRobs	134998	-0.37503	0.04632	-0.01803	0.00059	-0.01917	-0.01688	0.04665	0.21598
InHRdir1	134998	-0.35973	0.03968	-0.00273	0.00054	-0.00380	-0.00167	0.03968	0.19920
InHRdir2	134998	-0.38353	0.05323	-0.02653	0.00063	-0.02776	-0.02530	0.05394	0.23224
InHRindir5	134998	-0.36267	0.03361	-0.00567	0.00050	-0.00665	-0.00470	0.03365	0.18343
InHRindir6	134998	-0.32214	0.02493	0.03486	0.00043	0.03402	0.03570	0.02614	0.16169
InHRindir7	134998	-0.36026	0.03146	-0.00326	0.00048	-0.00421	-0.00232	0.03148	0.17741
<i>Summary Statistics and Bias for Methods 1 through 9 with HR true = 0.9 [log(HR true) = -0.105]:</i>									
InHRobs	134999	-0.11029	0.04365	-0.00529	0.00057	-0.00641	-0.00418	0.04368	0.20900
InHRdir1	134999	-0.10726	0.03961	-0.00226	0.00054	-0.00332	-0.00120	0.03961	0.19903
InHRdir2	134999	-0.11143	0.04607	-0.00643	0.00058	-0.00758	-0.00529	0.04611	0.21473
InHRindir5	134999	-0.17347	0.02427	-0.06847	0.00042	-0.06930	-0.06764	0.02896	0.17017
InHRindir6	134999	-0.15361	0.01799	-0.04861	0.00037	-0.04933	-0.04790	0.02035	0.14267
InHRindir7	134999	-0.17230	0.02308	-0.06730	0.00041	-0.06811	-0.06649	0.02761	0.16617

Tab. 2: Über alle Szenarien gemittelte Punktschätzer für  $\log(HR)$  gemäß Cox-Regression und Methoden 1, 2, 7, 8 und 9 („Theta“) mit dazugehöriger mittlerer Abweichung („Bias“) inkl. Standardfehler („MCSE“) vom jeweiligen  $\log(HR_{true})$  (vgl. kursive Zwischenüberschriften).

### Gepoolte $\log(HR)$ -Schätzer pro Hazard Ratio

Abbildung 5 visualisiert die Verteilung der  $\log(HR)$ -Schätzer, stratifiziert nach den vorgegebenen  $HR_{true} = \{0.1, 0.3, 0.5, 0.7, 0.9\}$ , anhand jeweils um  $\log(HR_{true})$  zentrierter Dichtekurven:

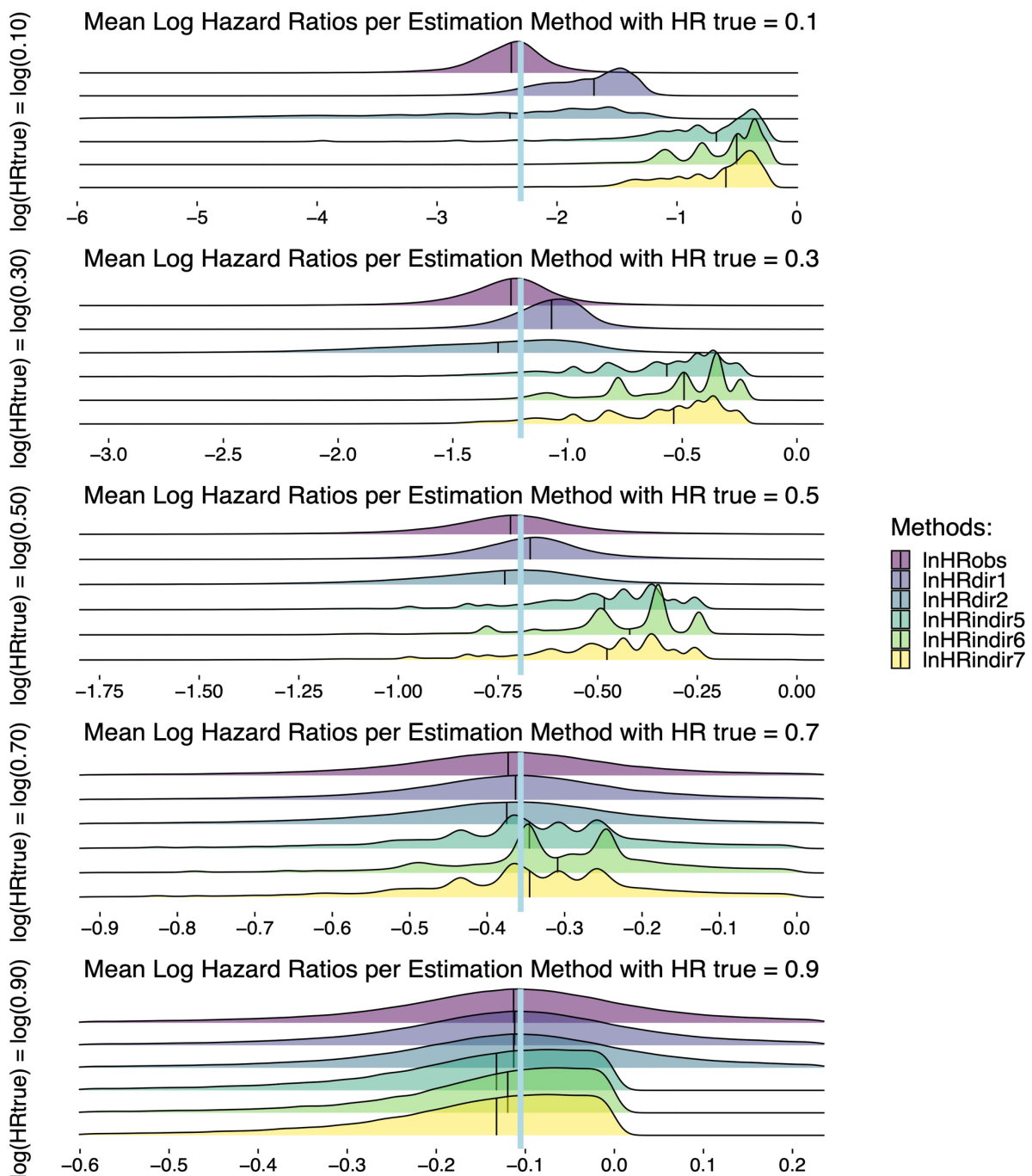


Abb. 4: Ridgeplots der  $\log(HR)$ -Punktschätzer (Formeln 4-8, Methoden 1-2,7-9, gestapelte Dichtekurven mit Median). Die hellblauen Vertikalen markieren die vorgegebenen  $\log(HR_{true})$ .



### Gepoolter Bias der $\log(HR)$ -Schätzer pro Hazard Ratio

Abbildung 6 visualisiert die mittleren Abweichungen der  $\log(HR)$ -Schätzer von den jeweiligen  $\log(HR_{true})$ -Werten anhand jeweils um 0 zentrierter Fehlerbalken, analog stratifiziert nach den vorgegebenen  $HR_{true} = \{0.1, 0.3, 0.5, 0.7, 0.9\}$ :

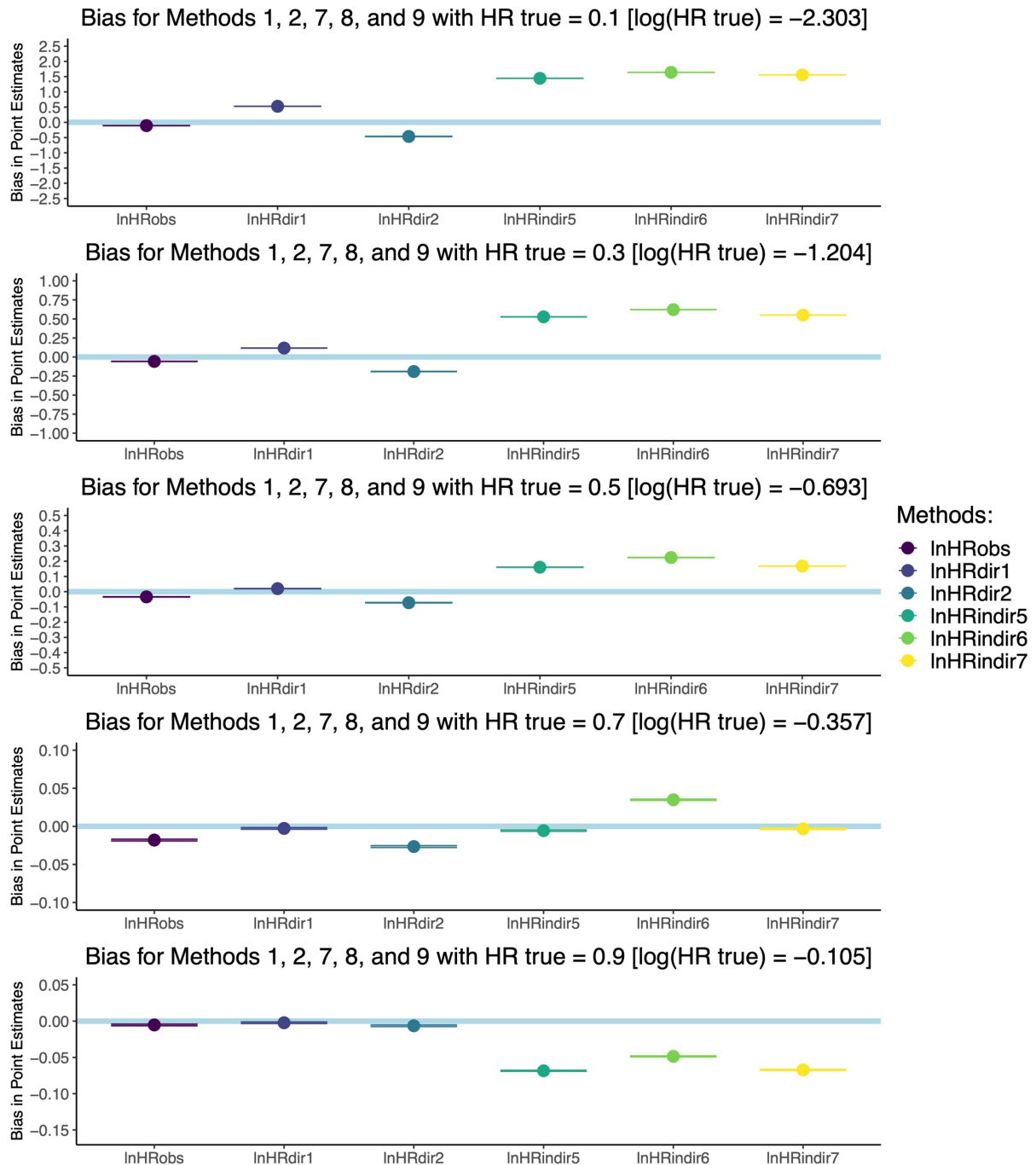


Abb. 5: Fehlerbalkendiagramme der absoluten mittleren Abweichungen der  $\log(HR)$ -Punktschätzer (Formeln 4-8, Methoden 1-2,7-9, x-Achse). Die hellblauen Horizontalen markieren einen absoluten Bias von 0, entsprechend unverzerrten Schätzern. Der Beträge der Abweichungen nehmen mit steigendem  $\log(HR_{true})$  ab, weshalb die Skalierung der y-Achsen feiner wird.

*Gepoolte log(HR)-Schätzer und gepoolter Bias pro Hazard Ratio unter faktorieller Variation der Designparameter „Allocation Ratio“, „Baseline Hazard“, „Censoring Distribution“ und „Trial Size“*

Nach globaler Anwendung der Funktion *calculate\_performances()* auf den lediglich nach den vorgegebenen  $HR_{true} = \{0.1, 0.3, 0.5, 0.7, 0.9\}$  stratifizierten Ergebnisdatensatz, wurde die Auswertung im nächsten Schritt um eine weitere Subgruppierung der stratifizierten Datensätze anhand der faktoriell variierten Designparameter erweitert. Die Abbildungen 7 bis 16 visualisieren nachfolgend jeweils paarweise anhand von Fehlerbalkendiagrammen zum einen und anhand von gestapelten Dichtekurven zum anderen den Einfluss verschiedener Allokationsraten (die jeweiligen Paneele sind mit „A“ für „Allocation Ratio“ gekennzeichnet), verschiedener Basisausfallraten (die jeweiligen Paneele sind mit „B“ für „Baseline Hazard“ gekennzeichnet), verschiedener Zensierungsraten (die jeweiligen Paneele sind mit „C“ für „Censoring Distribution“ gekennzeichnet) und verschiedener Gesamtfallzahlen (die jeweiligen Paneele sind mit „S“ für „Trial Size“ gekennzeichnet). Die zugrundeliegenden Daten sind in Analogie zu Tabelle 2 in den Tabellen 4 bis 23 im Anhang abgedruckt.

**Bias und Verteilung der Punktschätzer unter faktorieller Variation der Designparameter „Allocation Ratio“, „Baseline Hazard“, „Censoring Distribution“ und „Trial Size“ bei Hazard Ratio von 0.1:**

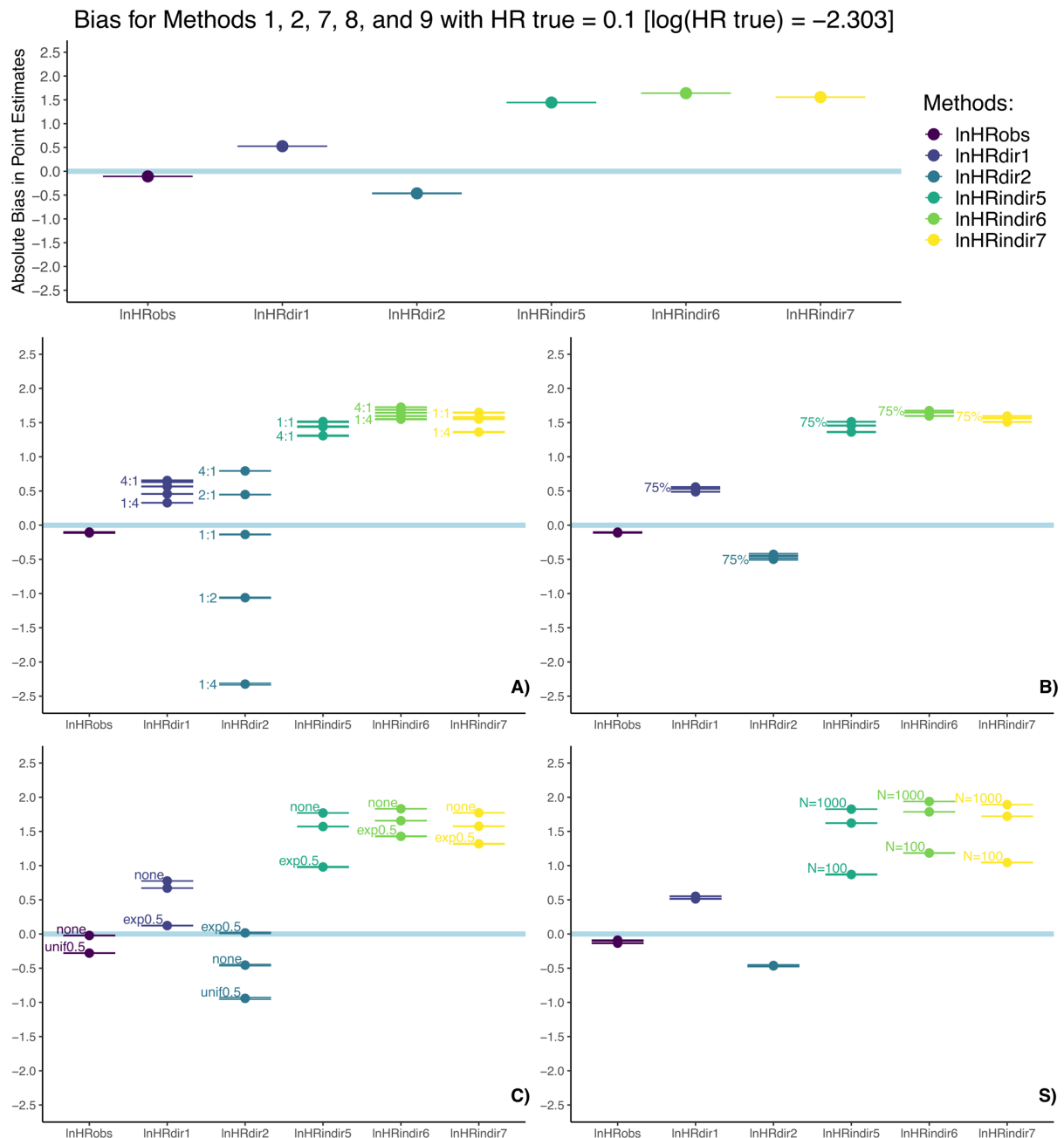


Abb. 6: Abweichungen der  $\log(\text{HR})$ -Punktschätzer bei vorgegebenem Hazard Ratio von 0.1; gepoolt (oben) und unter Veränderung der Allokationsrate (A), Basisausfallrate (B), Zensierungsrate (C) und Gesamtfallzahl (S). Die auf berichteten P-Werten basierenden Methoden 7 bis 9 (vgl. Formeln 6 bis 8) weisen den höchsten Bias auf, denn höhere Fallzahlen und kleinere P-Werte verzerren die zugrundeliegenden Schätzer stark (S). Unbalancierte Allokationsraten (A) hingegen verzerren insbesondere die Schätzmethode 2 (vgl. Formel 5). Zensierungen beeinflussen alle Schätzer (C), haben jedoch den geringsten Einfluss auf die Cox-Regression (InHRobs). Die hellblauen Referenzlinien markieren jeweils einen Bias von 0.

– Ergebnisse –

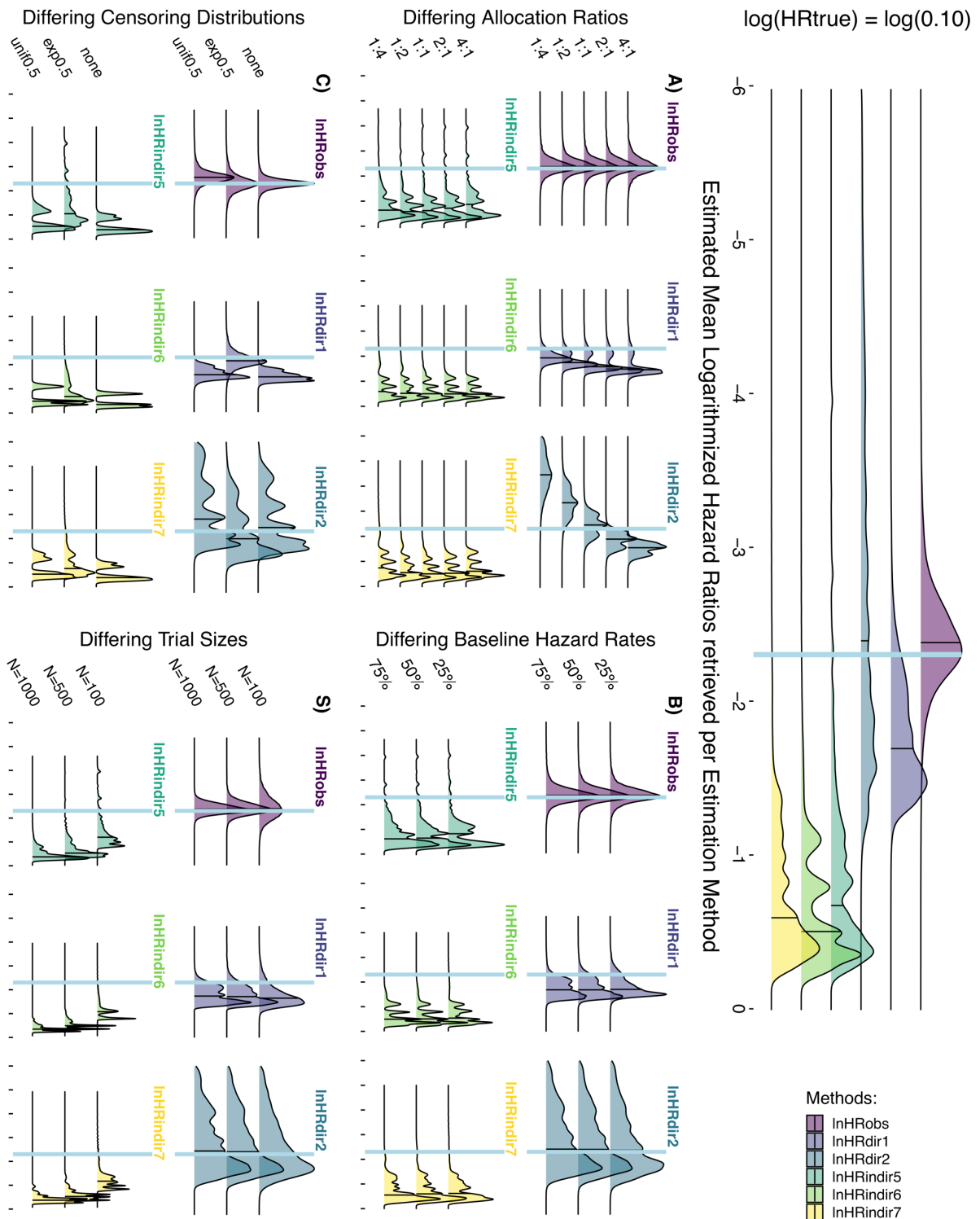


Abb. 7: Ridgeplots (gestapelte Dichtekurven mit Kennzeichnung des Median) der Verteilungen der  $\log(HR)$ -Punktschätzer bei vorgegebenem Hazard Ratio von 0.1; gepoolt (rechts) und unter Veränderung von Allokationsrate (A), Basisausfallrate (B), Zensierungsrate (C) und Gesamtfallzahl (S). Die hellblauen Referenzlinien markieren das vorgegebenen  $\log(HR_{true})$ . In Übereinstimmung mit Abb. 6 liegen die Methoden 7 bis 9 (InHRindir5 bis 7) am weitesten vom wahren Wert entfernt und wird die Schätzmethode 2 durch unbalanzierte Allokationsraten bidirektional verzerrt, was eine breite Streuung ihrer gepoolten Dichtekurve bedingt (InHRdir2).

**Bias und Verteilung der Punktschätzer unter faktorieller Variation der Designparameter „Allocation Ratio“, „Baseline Hazard“, „Censoring Distribution“ und „Trial Size“ bei Hazard Ratio von 0.3:**

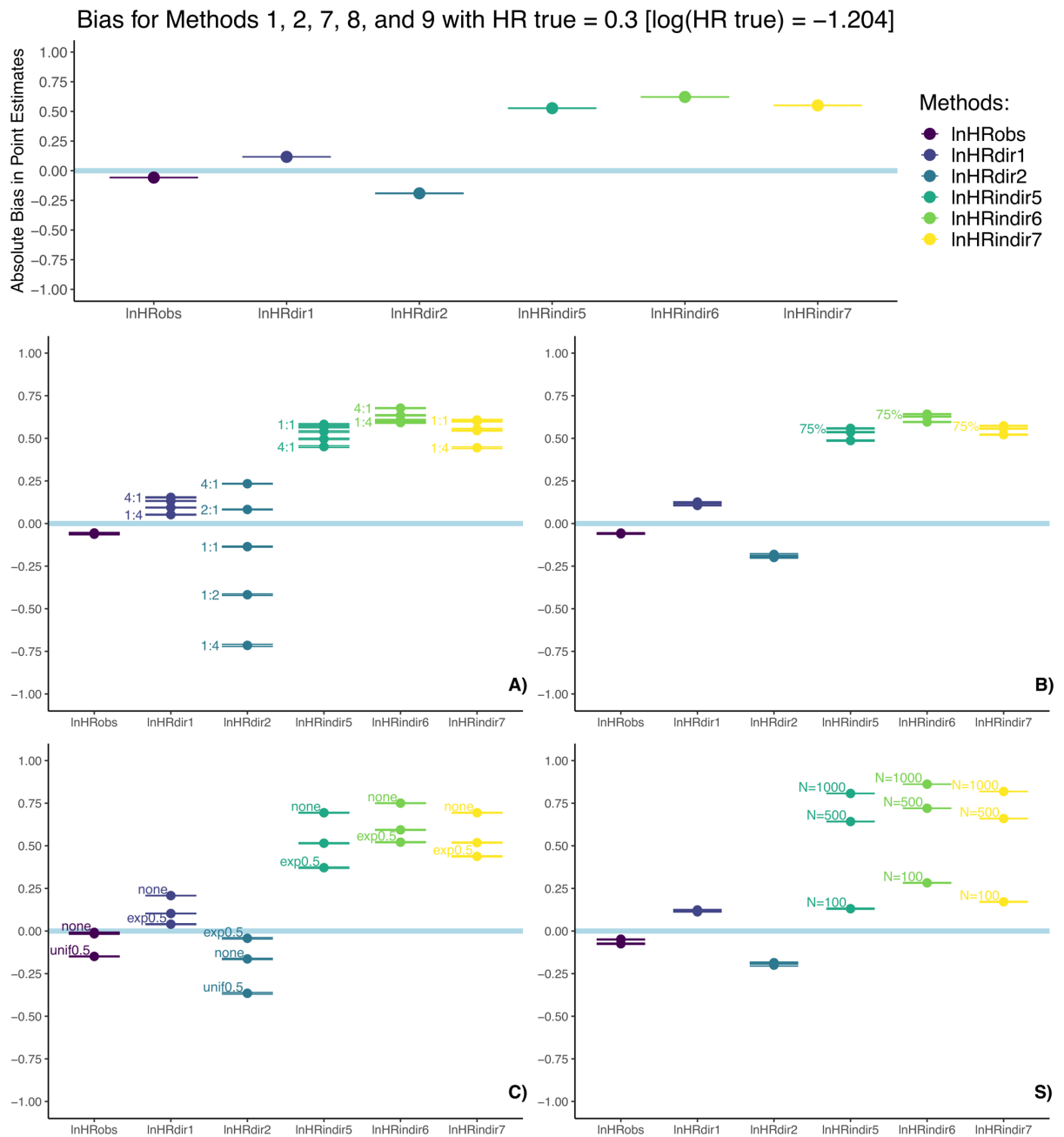


Abb. 8: Abweichungen der log(HR)-Punktschätzer bei vorgegebenem Hazard Ratio von 0.3; gepoolt (oben) und unter Veränderung der Allokationsrate (A), Basisausfallrate (B), Zensurierungsrate (C) und Gesamtfallzahl (S). Die auf berichteten P-Werten basierenden Methoden 7 bis 9 (vgl. Formeln 6 bis 8) weisen wiederum bei hohen Fallzahlen und kleinen P-Werten starke Verzerrungen auf (S). Unbalancierte Allokationsraten, insbesondere zugunsten der Experimentalgruppe (A), verzerren dagegen die Methode 2 (Formel 5). Zensurierungen beeinflussen alle Schätzer (C), haben jedoch einen geringeren Einfluss auf Schätzer der Cox-Regression und Methode 1. Die hellblauen Referenzlinien markieren jeweils einen Bias von 0.

– Ergebnisse –

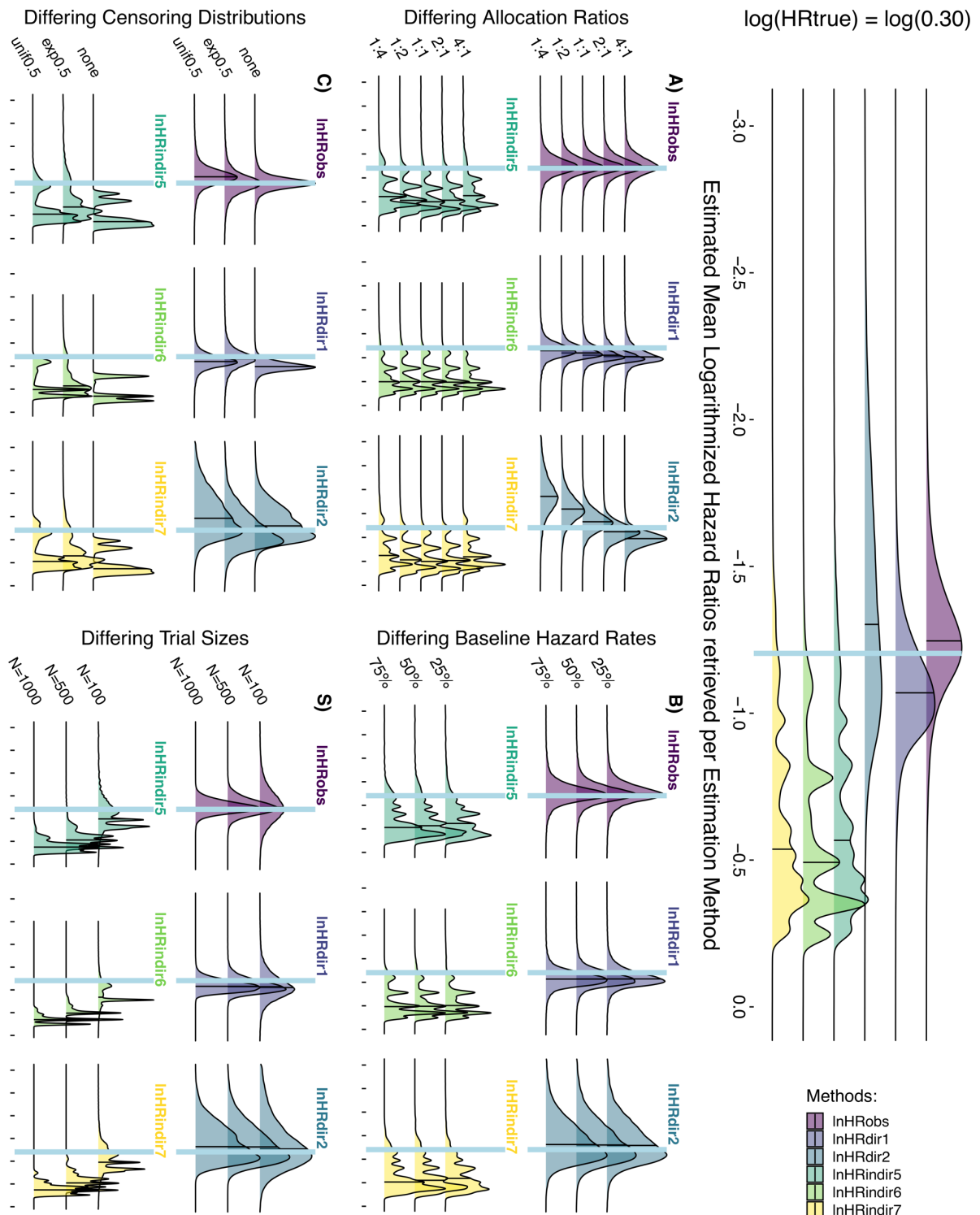


Abb. 9: Ridgeplots der Verteilungen der  $\log(HR)$ -Schätzer mit Kennzeichnung des Median bei vorgegebenem Hazard Ratio von 0.3; gepoolt (rechts) und unter Veränderung von Allokationsrate (A), Basisausfallrate (B), Zensierungsrate (C) und Gesamtfallzahl (S). Die hellblauen Referenzlinien markieren das vorgegebene  $\log(HR_{true})$ . In Übereinstimmung mit Abb. 8 liegen die Schätzmethoden 7 bis 9 (InHRindir5 bis 7) bei Fallzahlen über 100 (S) und die Schätzmethode 2 bei unbalancierten Allokationsraten (A) weit vom wahren Wert entfernt (InHRdir2).

**Bias und Verteilung der Punktschätzer unter faktorieller Variation der Designparameter „Allocation Ratio“, „Baseline Hazard“, „Censoring Distribution“ und „Trial Size“ bei Hazard Ratio von 0.5:**

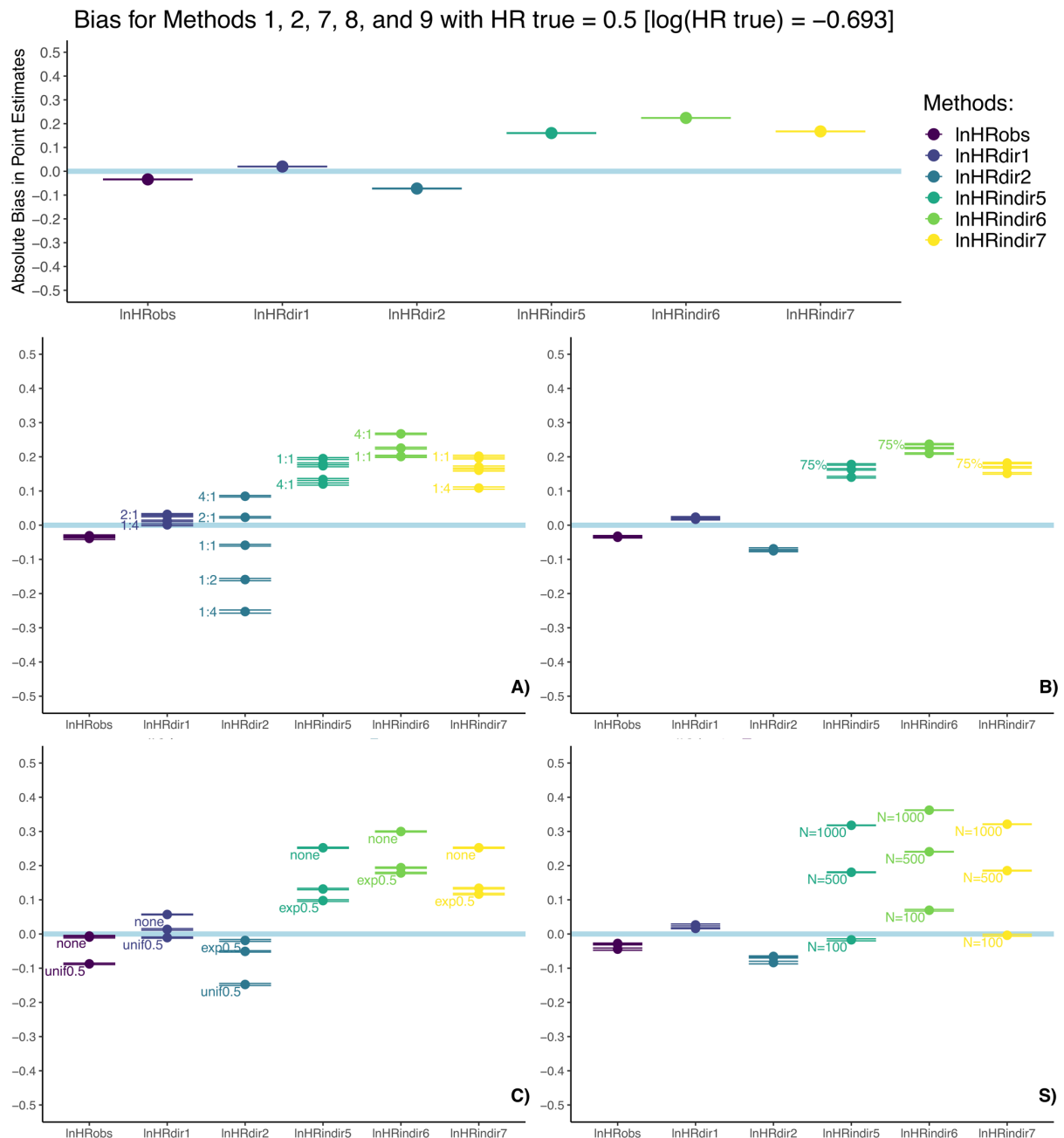


Abb. 10: Abweichungen der  $\log(\text{HR})$ -Punktschätzer bei vorgegebenem Hazard Ratio von 0.5; gepoolt (oben) und unter Veränderung der Allokationsrate (A), Basisausfallrate (B), Zensierungsrate (C) und Gesamtfallzahl (S). Nachwievor weisen die Methoden 7 bis 9 (Formeln 6 bis 8) bei Gesamtfallzahlen von 1000 und entsprechend kleinen P-Werten deutliche Verzerrungen auf (S). Die Schätzmethode 2 (Formel 5) wird weiterhin von extremen Allokationsimbilanzen zugunsten der Experimentalgruppe deutlich verzerrt (A). Zensierungen beeinflussen weiterhin alle Schätzer (C). Die hellblauen Referenzlinien markieren jeweils einen Bias von 0.

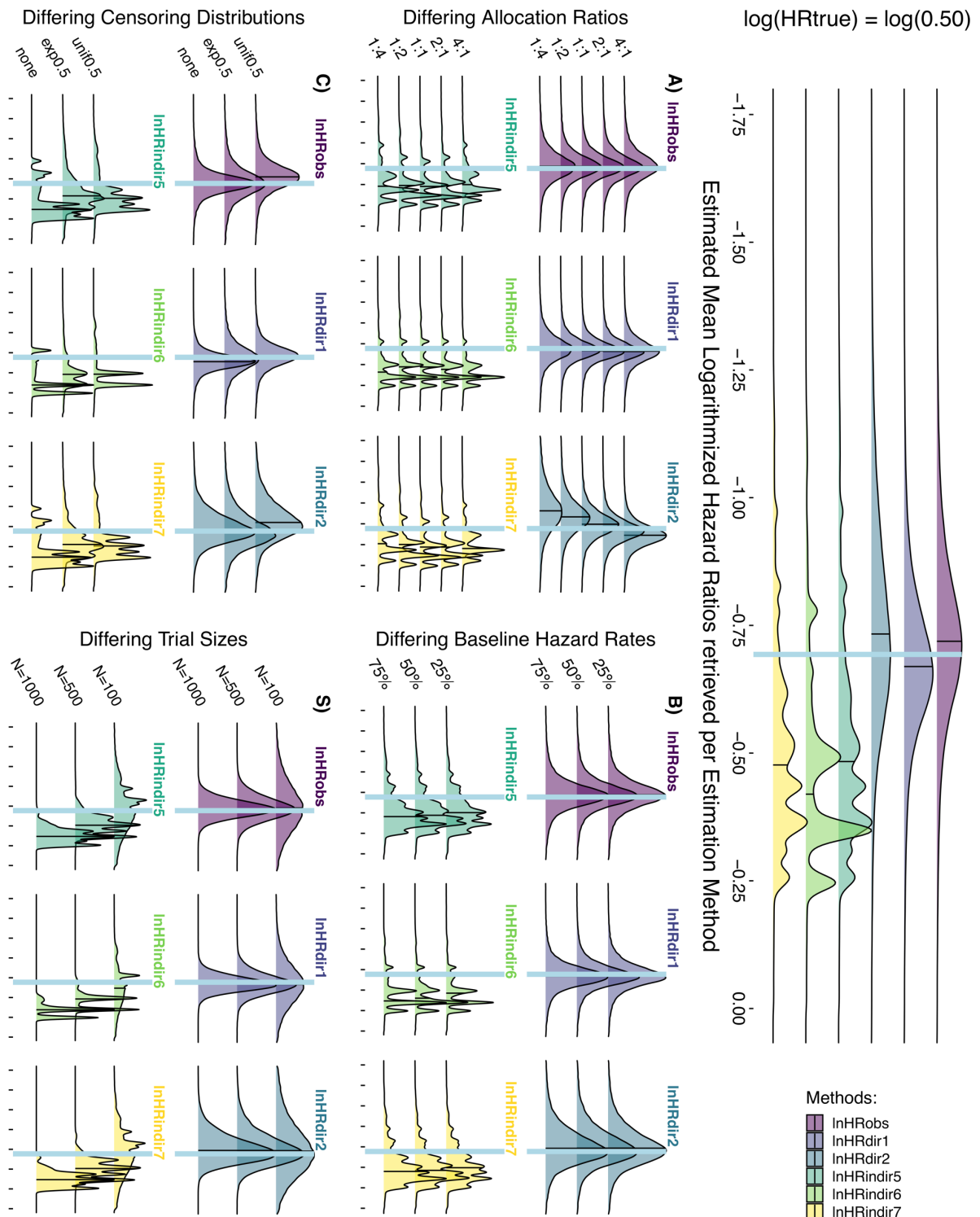


Abb. 11: Ridgeplots der Verteilungen der  $\log(HR)$ -Schätzer mit Kennzeichnung des Median bei vorgegebenem Hazard Ratio von 0.5; gepoolt (rechts) und unter Veränderung von Allokationsrate (A), Basisausfallrate (B), Zensierungsrate (C) und Gesamtfallzahl (S). Die hellblauen Referenzlinien markieren das vorgegebenen  $\log(HR_{true})$ . In Übereinstimmung mit Abb. 10 liegen die Verteilungen der Schätzwerte nunmehr zum einen insgesamt weniger weit von den wahren Werten entfernt und zum anderen mit ähnlichen Streubreiten. Die Methoden 7 bis 9 (InHRindir5 bis 7) weichen bei hohen Fallzahlen (S) dennoch deutlich ab.



**Bias und Verteilung der Punktschätzer unter faktorieller Variation der Designparameter „Allocation Ratio“, „Baseline Hazard“, „Censoring Distribution“ und „Trial Size“ bei Hazard Ratio von 0.7:**

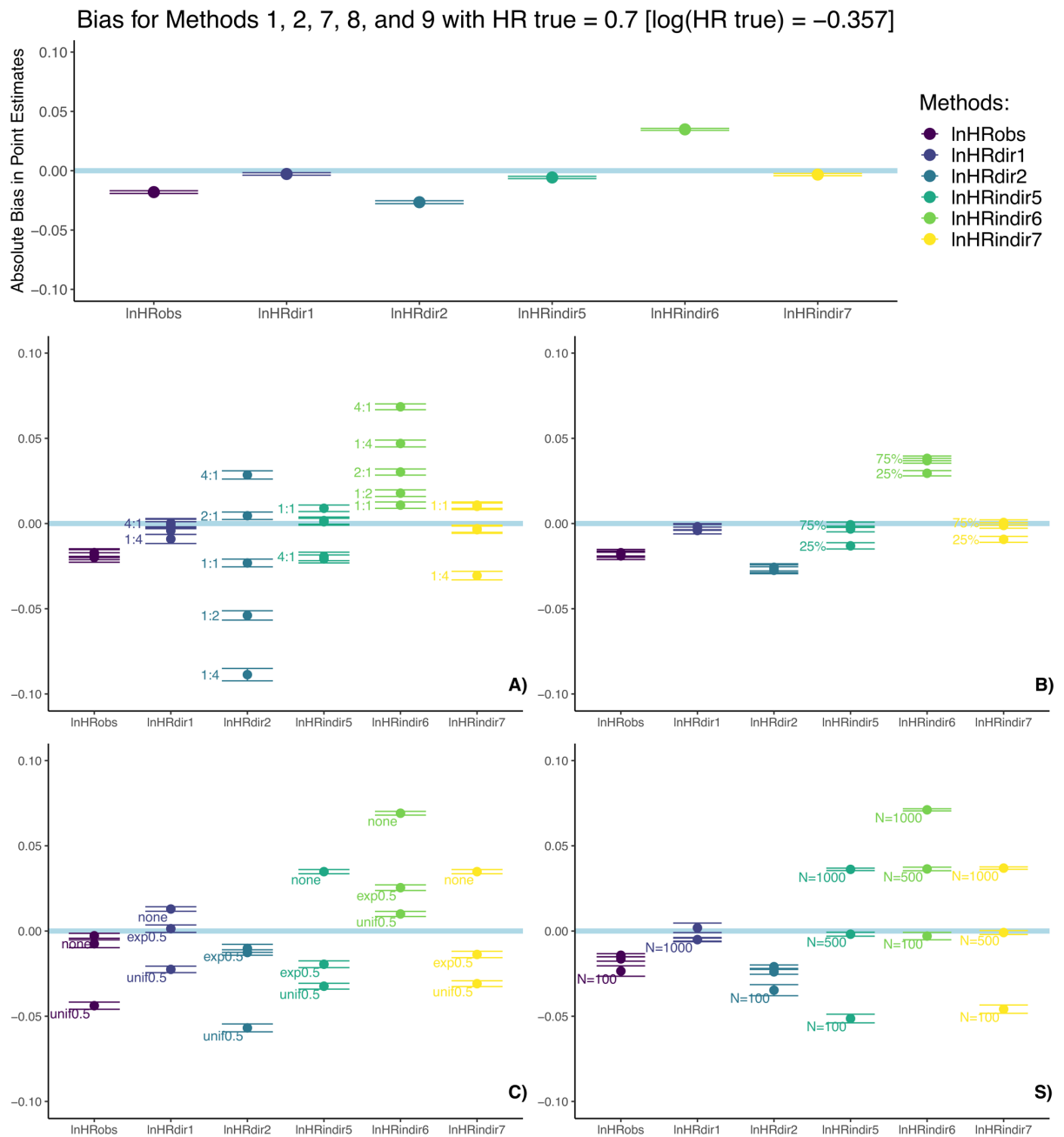


Abb. 12: Abweichungen der  $\log(\text{HR})$ -Punktschätzer bei vorgegebenem Hazard Ratio von 0.7; gepoolt (oben) und unter Veränderung der Allokationsrate (A), Basisausfallrate (B), Zensurierungsrate (C) und Gesamtfallzahl (S). Die Beträge der Abweichungen sind nun deutlich geringer. Die Methoden 2 (Formel 5), und 8 (Formel 7) sind global betrachtet am stärksten verzerrt, was insbesondere auf Szenarien mit Allokationsungleichgewichten zurückzuführen ist (A). Verschiedene Zensurierungsmuster (C) beeinflussen alle Schätzer in verschiedene Richtungen. Hohe Fallzahlen bedingen eine positive Verzerrung des Punktschätzers 8, niedrige Fallzahlen hingegen eine negative Verzerrung der Punktschätzers 7 und 9 (S).

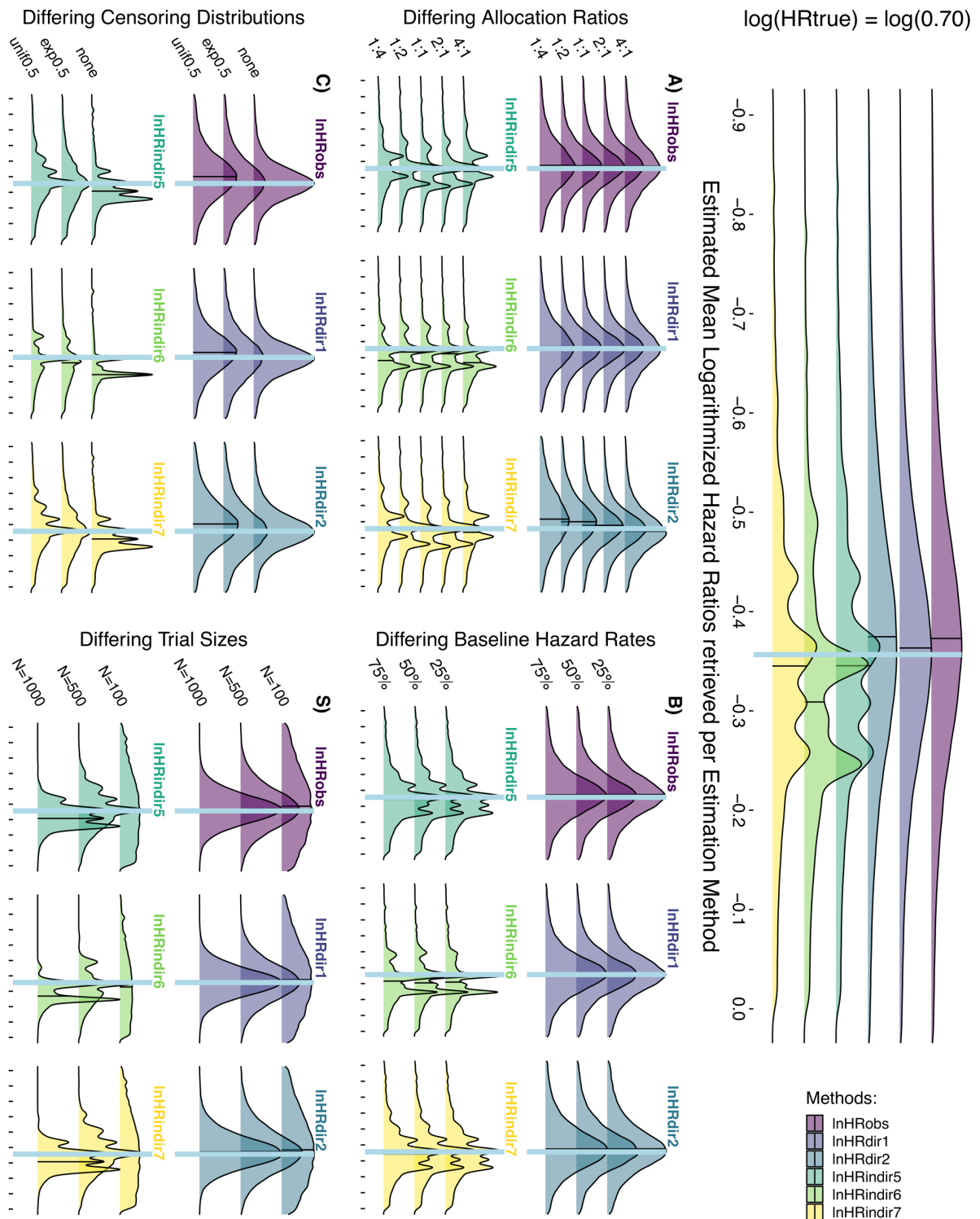


Abb. 13: Ridgeplots der Verteilungen der  $\log(HR)$ -Schätzer mit Kennzeichnung des Median bei vorgegebenem Hazard Ratio von 0.7; gepoolt (rechts) und unter Veränderung von Allokationsrate (A), Basisausfallrate (B), Zensierungsrate (C) und Gesamtfallzahl (S). Die hellblauen Referenzlinien markieren das vorgegebenen  $\log(HR_{true})$ . In Übereinstimmung mit Abb. 12 liegen die Verteilungen der Schätzwerte nah am wahren  $\log(HR_{true})$ , mit Ausnahme der Schätzmethode 8. Diese weist bei extremen Allokationsungleichgewichten (A) und hohen Fallzahlen (S) deutliche positive Kurvenverschiebungen auf (InHRdir6).

**Bias und Verteilung der Punktschätzer unter faktorieller Variation der Designparameter „Allocation Ratio“, „Baseline Hazard“, „Censoring Distribution“ und „Trial Size“ bei Hazard Ratio von 0.9:**

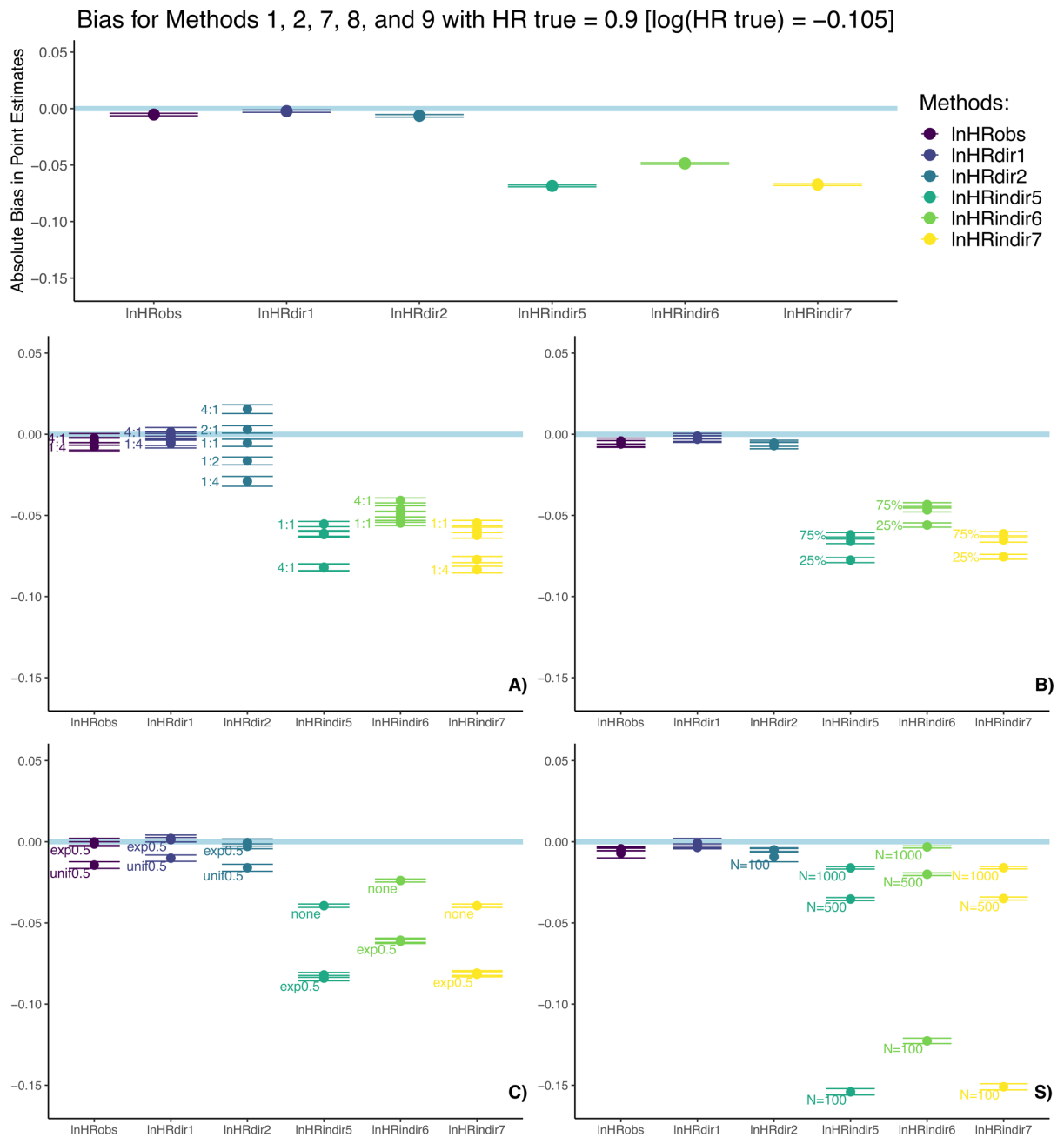


Abb. 14: Abweichungen der  $\log(HR)$ -Punktschätzer bei vorgegebenem Hazard Ratio von 0.9; gepoolt (oben) und unter Veränderung der Allokationsrate (A), Basisausfallrate (B), Zensurierungsrate (C) und Gesamtfallzahl (S). Die Beträge der Abweichungen sind nachwievor gering. Leidlich die auf berichteten P-Werten basierenden Methoden 7 bis 9 (vgl. Formeln 6 bis 8) sind mit einer nunmehr negativen Abweichung assoziiert, was insbesondere auf Szenarien mit Allokationsungleichgewichten (A), geringeren Basisausfallraten (B), sowie niedrigeren Ereignisinzidenzen aufgrund von etwaigen Zensurierungen (C) und geringeren Fallzahlen (S) zurückzuführen ist.

– Ergebnisse –

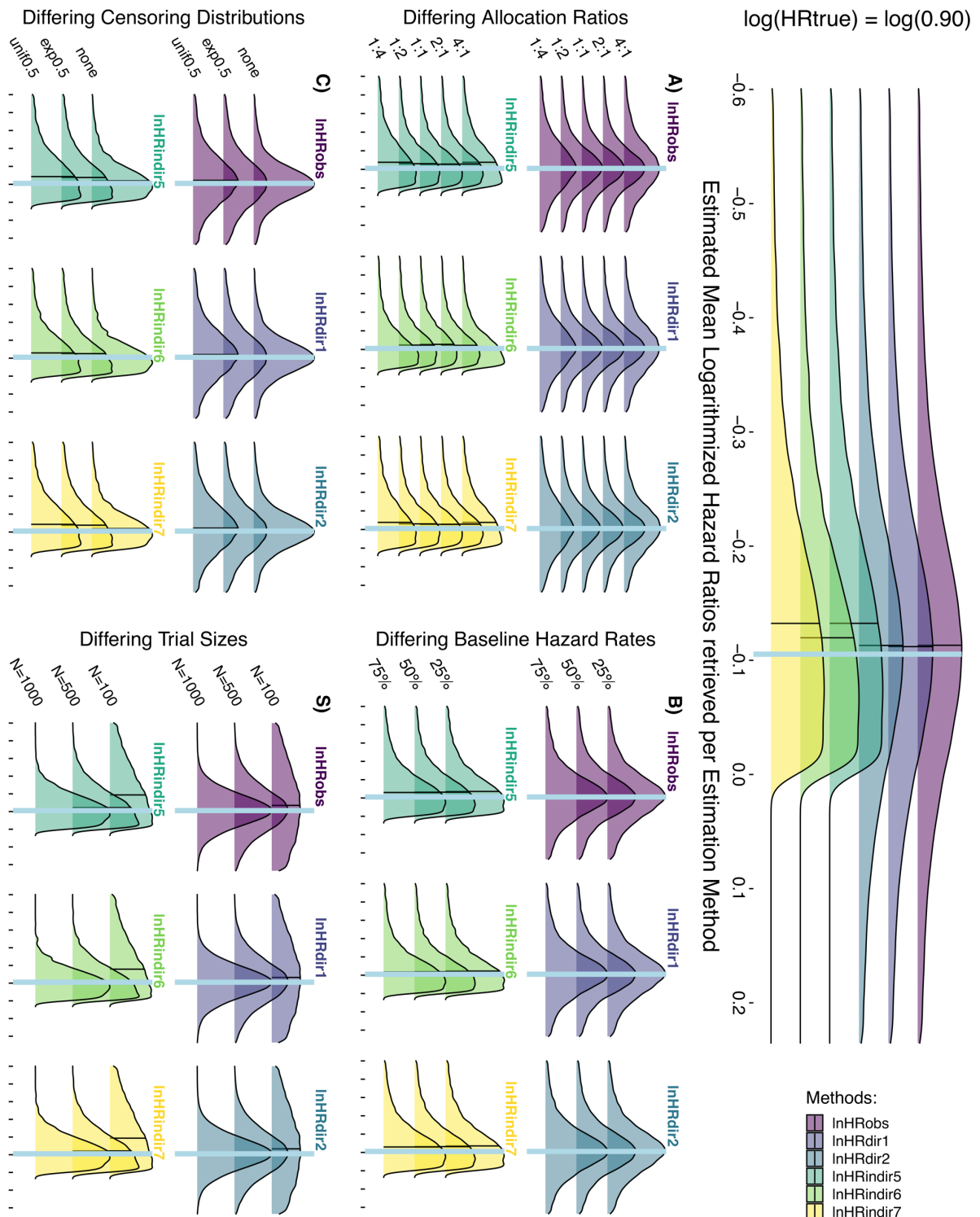


Abb. 15: Ridgeplots der Verteilungen der  $\log(HR)$ -Schätzer mit Kennzeichnung des Median bei vorgegebenem Hazard Ratio von 0.9; gepoolt (rechts) und unter Veränderung von Allokationsrate (A), Basisausfallrate (B), Zensierungsrate (C) und Gesamtfallzahl (S). Die hellblauen Referenzlinien markieren das vorgegebenen  $\log(HR_{true})$ . In Übereinstimmung mit Abb. 14 liegen die Verteilungen aller Schätzwerte nah am wahren Wert, mit Ausnahme der Schätzmethode 7 bis 9 (InHRindir5 bis 7) bei niedrigen Fallzahlen (S).

*Performance measures der var(log(HR))-Punktschätzer*

Tabelle 3 listet die Ergebnisse der Funktion *calculate\_variances()* angewendet auf die Ergebnisvektoren der Cox-Regression (Spalte „lnVobs“) und der Methoden 1 bis 6 (Spalten „lnVdir1“, „lnVdir2“, „lnVindir1“, „lnVindir2“, „lnVindir3“ und „lnVindir4“). Die Spalte „Methods“ bezeichnet die jeweils analysierte Auswertungsmethode und die Spalte „Iterations“ die Anzahl der zugrundeliegenden Simulationen. Die Spalte „Theta“ enthält die arithmetischen Mittelwerte der geschätzten *var(log(HR))* und die Spalte „Variance“ die dazugehörigen empirischen Varianzen, wobei pro Methode und vorgegebenem Hazard Ratio über alle Szenarien mit jeweils  $n_{sim} = 1000$  pro Szenario gepoolt wurde (s. „Material und Methoden: Szenarien der vorgelegten Simulationsstudie“). Die Spalten „MCSE“ und „Lower“ bzw. „Upper“ enthalten die dazugehörigen Standardfehler und ihre 95% Konfidenzintervalle.

Pooled Variance Estimators for Methods 1 through 6

<b>Method</b>	<b>Iterations</b>	<b>Theta</b>	<b>Variance</b>	<b>MCSE</b>	<b>Lower</b>	<b>Upper</b>
lnVobs	673642	0.06356	0.01104	0.00013	0.06331	0.06381
lnVdir1	673642	0.05507	0.00723	0.00010	0.05487	0.05527
lnVdir2	673642	0.05914	0.00823	0.00011	0.05893	0.05936
lnVindir1	673642	0.06252	0.00946	0.00012	0.06229	0.06276
lnVindir2	673642	0.05182	0.00773	0.00011	0.05161	0.05203
lnVindir3	673642	0.03479	0.00190	0.00005	0.03468	0.03489
lnVindir4	673642	0.04452	0.00346	0.00007	0.04438	0.04466

Tab. 3: Über alle Szenarien gemittelte Punktschätzer für *var(log(HR))* gemäß Cox-Regression und Methoden 1 bis 6 („Theta“) mit dazugehöriger empirischer Varianz („Variance“) inkl. Standardfehler („MCSE“) und 95% Konfidenzintervall.

*Relative Studiengewichte basierend auf verschiedenen var(log(HR))-Schätzern unter faktorieller Variation aller Designparameter*

Nach der globalen Anwendung von *calculate\_variances()* auf den unstratifizierten Ergebnisdatensatz, wurde die Auswertung im nächsten Schritt anhand aller faktoriell variierten Designparameter durchgeführt. Mithilfe der inversen Varianz-Methode wurden sodann die relativen Gewichte etwaiger Überlebenszeitstudien bestimmt,

deren Inklusion in eine etwaige Metaanalyse Datenextraktionen mittels der Methoden 1 bis 6 (bzw. 1 bis 9) notwendig machen würde (s. „Material und Methoden: Performance measures 2: Relative Studiengewichte“). Abbildung 17 visualisiert die resultierenden relativen Studiengewichte im Vergleich zur Verwendung der Cox-Regression:

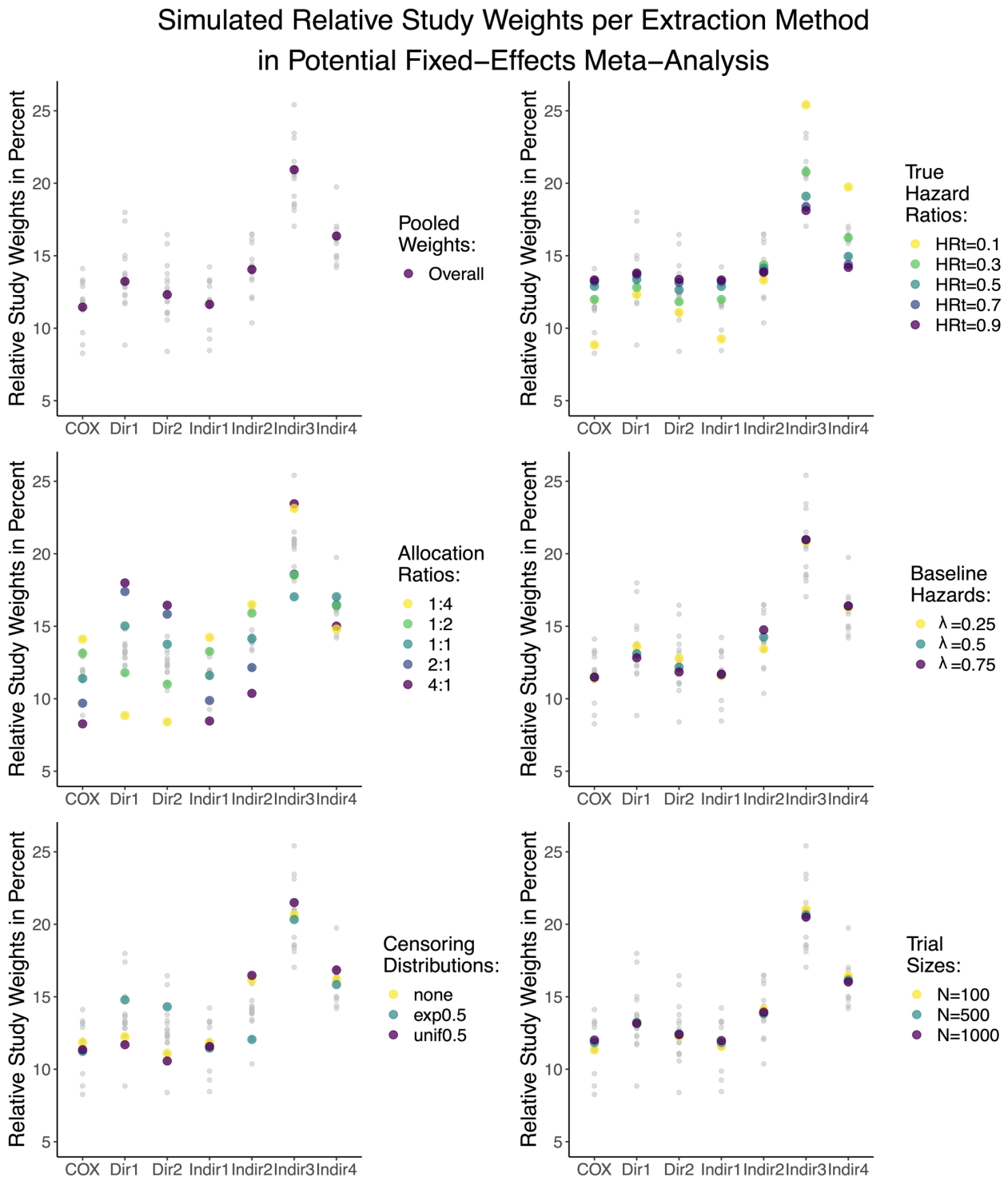


Abb. 16: Scatterplots simulierter relativer Studiengewichte ermittelt mithilfe der inversen Varianz-Methode auf Grundlage der  $\text{var}(\log(\text{HR}))$ -Schätzer der Methoden 1 bis 6 (Formeln 9 bis 14).

Die  $\text{var}(\log(HR))$ -Schätzer der Methoden 7 bis 9 entsprechen dabei denen der Methoden 4 bis 6 und sind daher nicht mit abgebildet ( $\text{InVdir5} \triangleq \text{InVdir2}$ ,  $\text{InVdir6} \triangleq \text{InVdir3}$  und  $\text{InVdir7} \triangleq \text{InVdir4}$ ). Die Methoden 5 und 6 bzw. 8 und 9 schätzen deutlich geringere Varianzen im Vergleich zur Cox-Regression. Eingeschlossene Studien basierend auf anhand dieser Methoden extrahierter Daten (Formeln 13 und 14) erhalten im Verhältnis deutlich höhere Gewichte.

Wie auch die Ergebnisse der Subgruppenanalysen mithilfe der `calculate_performances()`-Funktion in Form der Tabellen 4 bis 23, sind die zugrundeliegenden Daten (gemäß Anwendung der `calculate_variances()`-Funktion) als Tabellen 24 bis 28 im Anhang zu finden.

## **Diskussion und Fazit**

Im Folgenden wird die Motivation für die vorliegende Simulationsstudie nochmals dargelegt und die zugrundeliegende Fragestellung hergeleitet sowie anhand der bis dato verfügbaren Literatur zu eben jener Fragestellung a priori zu erwartende Trends beleuchtet. Sodann werden die Ergebnisse der vorliegenden Simulationsstudie zusammengefasst und im Anschluss im Bezug zu den Ihnen zugrundeliegenden Berechnungsmethoden interpretiert und bewertet. Als Fazit wird abschließend eine szenarienbasierte Anwendungsempfehlung der untersuchten Methoden unter Berücksichtigung ihrer Performancegrenzen auf Grundlage der Ergebnisse formuliert. Im Nachgang werden wichtige Limitationen der vorliegenden Simulationsstudie erläutert und Ansatzpunkte für weiterführende Untersuchungen aufgezeigt.

### *Rekapitulation der Ausgangslage und Motivation der Fragestellung*

Die Erkenntnissynthese publizierter biomedizinischer Studien anhand systematischer Reviews und Metanalysen ist unverzichtbarer Bestandteil der Evidenz-basierten Medizin.<sup>3,5,13</sup> Allerdings werden insbesondere im Kontext der Überlebenszeitanalyse die dafür benötigten Kenngrößen, d.h. Hazard Ratio und dazugehörige Streuung, in der biomedizinischen Fachliteratur häufig nicht ausreichend berichtet.<sup>10–12</sup> Um etwaige Studien dennoch berücksichtigen zu können, wurden zahlreiche Schätzmethoden zur Approximation der Hazard Ratios und ihrer dazugehörigen Varianzen aus aggregierten Daten veröffentlicht.<sup>9,16–19</sup> Allerdings warnen die jeweiligen Autoren vor potentiell zunehmender Ungenauigkeit der Schätzmethoden, gerade bei kleineren Fallzahlen, niedrigeren Eventraten, starken Therapieeffekten und Gruppenimbancen, ergo im Falle von Szenarien, welche in der klinischen Realität nicht selten anzutreffen sind.<sup>9,19–21</sup> Eine systematische Untersuchung der Genauigkeit der Schätzmethoden fehlte jedoch bislang, genauso wie daraus potentiell ableitbare Anwendungsempfehlungen. Die vorliegende Arbeit hatte daher zum Ziel den potentiellen Einfluss verschiedener Designparameter auf die Performance der vorgeschlagenen Schätzmethoden zu quantifizieren und eine Empfehlung zum situationsspezifischen Umgang mit den Schätzmethoden abzuleiten.



### *Anhand bisher verfügbarer Literatur erwartbarer Ergebnisanteil*

#### **Bias der $\log(HR_{obs})$ -Schätzer im Cox-Modell**

Zunächst muss der im Rahmen der Simulationsstudie kalkulierte Bias der Cox-Regression diskutiert werden. Die berechneten  $\log(HR_{obs})$ -Werte wichen im Mittel zwischen -0.108 und -0.005 von den  $\log(HR_{true})$ -Werten ab. Die Schätzer der Cox-Regression wiesen also einen negativen Bias auf. Im Falle eines „wahren“ Hazard Ratios von 0.1 betrug der  $\log(HR_{obs})$ -Schätzer -2.411 entsprechend einem Hazard Ratio von 0.0897. Im Falle eines „wahren“ Hazard Ratios von 0.9 betrug der  $\log(HR_{obs})$ -Schätzer -0.110 entsprechend einem Hazard Ratio von 0.896. Die Abweichungen waren also stets negativ und größer im Falle größerer Therapieeffekte bzw. kleiner im Falle kleinerer Therapieeffekte. Diese geringfügigen Abweichungen waren jedoch aufgrund der Verwendung von (Rechts-) Zensierungen zu erwarten.

#### **Bias der approximativen $\log(HR)$ - und $\text{var}(\log(HR))$ -Schätzer**

Parmar und Kollegen warnten bereits vor potentiellen Problemen mit ungenau berichteten P-Werten und damit assoziierten Verzerrungen bei der Verwendung der selbigen zur Kalkulation von Therapieeffekten.<sup>9</sup> Somit war eine Anfälligkeit der Methoden 7 bis 9 für Rundungsfehler zu erwarten. Zudem hatten Parmar et al. relevante Verzerrungen der  $\log(HR)$ -Schätzmethode 2 im Falle niedriger Gesamtereigniszahlen beobachtet.<sup>9</sup> Dementsprechend wäre eine Anfälligkeit der Methode 2 für niedrigere Eventraten zu erwarten gewesen. Darüber hinaus wiesen die Autoren auf eine potentielle Fehlschätzung der Varianz mithilfe der Approximation  $\text{var}(HR) \approx \text{Gesamtereignisse}/4$  (Formel 13) in Szenarien mit großen Therapieeffekten und unbalanzierten Allokationsverhältnissen hin.<sup>9</sup> Somit war eine Anfälligkeit der Methoden 5 und 8 in solchen Szenarien zu erwarten. Jayne Tierney machte darüber hinaus in ihrer Publikation deutlich, dass Methoden 4 und 5 im Falle unbalanzierter Studiendesigns unangebracht seien, da die Varianz falsch geschätzt würde.<sup>19</sup> Von einer potentiellen Fehlgewichtung unbalanzierter Studien, deren  $\text{var}(\log(HR))$ -Werte mittels dieser Methoden geschätzt würden, war daher auszugehen. Zudem wiederholte sie die warnenden Worte Parmars bzgl. ungenau berichteter P-Werte.<sup>19</sup>

### *Zusammenfassung der Ergebnisse der vorliegenden Simulationsstudie*

In der vorliegenden Simulationsstudie wurde der Einfluss verschiedener Therapieeffektstärken, unbalanzierter Allokationsraten, unterschiedlicher Basisausfall- und Zensierungsraten, sowie der Gesamtfallzahl auf die Performance approximativer

$\log(HR)$ - und  $\text{var}(\log(HR))$ -Punktschätzer untersucht. Hierzu wurden exponentialverteilte rechtszensierte Überlebenszeitdaten mit 1000 Wiederholungen pro spezifischer faktoriell alterierter Kombination der vorgenannten Designparameter simuliert und diese mittels Cox-Regression versus vorgenannter Schätzmethoden analysiert. Dabei wurden folgende Beobachtung gemacht:

#### **Bias und MCSE der approximativen $\log(HR)$ -Schätzer**

Alle fünf untersuchten  $\log(HR)$ -Schätzmethoden wiesen einen Bias im Vergleich zu den jeweils vorgegebenen  $\log(HR_{true})$ -Werten auf, wobei der Betrag der Abweichung tendenziell mit abnehmendem Therapieeffekt ebenso abnahm. Die  $\log(HR)$ -Schätzmethode 1 (vgl. Formel 4) wies dabei im Mittel die geringsten Verzerrungen auf. Diese bewegten sich zwischen 0.526 und -0.002, was relativen Abweichungen von weniger als 25% entspricht (s. Tab. 2). Methode 2 (vgl. Formel 5) wies im Mittel die höchste Streuung bzw. geringste Präzision auf. Die breiteste Streuung der Schätzwerte zeigte sich dabei unter Variation der Allokationsverhältnisse mit Abweichungen von -34% bis +101% bei vorgegebenem Hazard Ratio von 0.1 und von -15% bis +28% bei vorgegebenem Hazard Ratio von 0.9 (s. Tab. 4 und 20 im Anhang). Die auf berichteten P-Werten basierenden Methoden 7 bis 9 (vgl. Formeln 6 bis 8) wiesen im Mittel die höchste gepoolte Verzerrung auf, nämlich -63%, -71%, -68% bei  $HR_{true}$  von 0.1 und 65%, 46%, 64% bei  $HR_{true}$  von 0.9. Entsprechend dem Zusammenhang zwischen Fallzahl bzw. Ereigniszahl und P-Wert, wiesen die Designparameter Gesamtfallzahl und Zensierungsrate einen großen Einfluss auf die Schätzmethoden 7 bis 9 auf. Die dazugehörigen MCSE lagen stets unter 0.005.

#### *Interpretation und Bewertung der Ergebnisse der vorliegenden Simulationsstudie*

##### **Beobachteter Bias der $\log(HR)$ -Schätzer der Cox-Regression**

Für den beobachteten Bias der  $\log(HR)$ -Schätzer der Cox-Regression bieten sich mehrere Erklärungsansätze an. Insbesondere die Reduktion der beobachteten Ereignisse durch Zensierungen führt unweigerlich zur Zunahme der Ungenauigkeit im Cox-Modell, was beispielsweise auch bei der Fallzahlplanung berücksichtigt werden muss.<sup>30</sup> In der vorliegenden Simulationsstudie erzeugte die Cox-Regression leicht negativ verzerrte  $\log(HR)$ -Schätzer, schätzte also etwas größere Therapieeffekte. Dieselbe Beobachtung wurde bereits von Persson und Khamis in einer 2005 publizierten Simulationsstudie beschrieben, die unter proportionalen Hazards einen

Anstieg des Bias der Cox-Methode mit zunehmendem Anteil der Zensierungen und abnehmenden Gesamtfallzahlen beobachtet.<sup>31</sup> Analoge Verzerrungstrends bestätigen sich auch im Vergleich der absoluten Abweichungen der InHRobs-Schätzer in Szenarien mit zusätzlichen Zensierungen (s. Tab. 6, 10, 14, 18, 22 im Anhang) und in Szenarien mit geringeren Fallzahlen (s. Tab. 7, 11, 15, 19, 23 im Anhang).

### **Beobachteter Bias der $\log(HR)$ -Schätzer der Methoden 1 bis 9**

Schätzmethode 1 für das Hazard Ratio entspricht der direkten Berechnung des Verhältnisses der Hazardraten beider Gruppen (Formel 4). Dementsprechend zeigten sich im Großteil der Szenarien nur sehr geringe Verzerrungen des Schätzers. Auffällig war jedoch, dass in Szenarien mit vorgegebenem Hazard Ratio von 0.1 zu geringe Therapieeffekte berechnet wurden. Hierfür kommen entweder ein relativ zu großer Zähler (Hazardrate der Experimentalgruppe) oder aber ein relativ zu kleiner Nenner (Hazardrate der Kontrollgruppe) bei der Berechnung infrage. Dass sich in Szenarien mit einem Allokationsverhältnis von 1:4 der Bias nahezu halbierte (s. Tab. 4 im Anhang), scheint ein Hinweis auf eine relativ zu geringe berechnete Hazardrate der Kontrollgruppe zu sein. Ein weiteres Indiz für eine relativ zu geringe berechnete Hazardrate der Kontrollgruppe bietet ein Blick in Tabelle 6 im Anhang, die verdeutlicht, dass sich der Bias insbesondere unter exponentialverteilten Zensierungen reduzierte. Dieses Zensierungsmuster bedingt eine höhere Rate an Zensierungen zu Beginn des Beobachtungszeitraumes und damit gerade in Szenarien mit großen Therapieeffekten eine niedrigere Rate an beobachteten Events in der Kontrollgruppe. Schätzmethode 2 für das Hazard Ratio basiert auf dem Verhältnis der Differenz der beobachteten und erwarteten Ereignisse in der Experimentalgruppe (O-E) und der Varianz der Log-Rang-Teststatistik. Der Zähler dieses Verhältnisses, die vorgenannte Differenz O-E, wird maßgeblich durch unausgeglichene Allokationsverhältnisse beeinflusst und führt zur Schätzung geringerer Therapieeffekte bei Imbalancen zugunsten der Kontrollgruppe sowie zur Schätzung größerer Therapieeffekte bei Imbalancen zugunsten der Experimentalgruppe (s. Abb. 6A, 8A, 10A, 12A, 14A und Tab. 4, 8, 12, 16, 20 im Anhang). Die jeweiligen Abweichungen der Punktschätzer der Methode 2 von den vorgegebenen  $\log(HR_{true})$ -Werten waren dagegen im Falle von ausgeglichenen Allokationsverhältnissen gering. Die Eventrate für sich genommen, hatte entgegen der Vermutung von Parmar keinen starken Einfluss.<sup>9</sup> Die Schätzmethoden 7, 8 und 9 stellen Abwandlungen der Methode 2 dar und unterscheiden sich in der Herleitung des Zählers, der bereits diskutierten Differenz O-E. Die Approximationen dieser

Differenzen setzen sich in den Methoden 7, 8 und 9 jeweils aus Produkten zusammen, in die die Z-Werte der berichteten P-Werte einfließen (s. Formeln 6, 7 und 8). Da diese entsprechen der APA-Konvention in der vorliegenden Simulationsstudie auf drei Nachkommastellen gerundet wurden, ergibt sich für diesen Faktor ein maximaler Wert von  $\Phi^{-1}(1 - 0.001/2) = 3.291$ . Durch so geartete Rundungsfehler fällt der Zähler in Formel 5 zu klein aus, was sich im positiven Bias der drei Schätzmethoden insbesondere in Szenarien mit großen Therapieeffekten niederschlägt (s. Tab. 2 oben). Der Informationsverlust durch Rundungsfehler im Vergleich zum exakten P-Wert nimmt mit höherer Gesamtfallzahl  $n_{\text{obs}}$  zu, weshalb der positive Bias im Vergleich von  $n = 100$  zu  $n = 1000$  stark zunimmt (s. Abb. 6S und Tab. 7 im Anhang). Der vorgenannte Effekt des Informationsverlusts durch Rundungsfehler und damit die Verzerrung der Schätzmethoden 7, 8 und 9 dreht sich in Szenarien mit geringen Therapieeffekten allerdings um, da die exakten P-Werte nahe bei 1 liegen und hier auf 1 aufgerundet werden. Dies macht sich dann wiederum als negativer Bias der drei Schätzmethoden insbesondere in Szenarien mit kleinen Therapieeffekten bemerkbar (s. Tab. 2 unten). Hier nimmt der Informationsverlust durch Rundungsfehler im Vergleich zum exakten P-Wert entsprechend mit geringerer Gesamtfallzahl  $n_{\text{obs}}$  zu, weshalb der negative Bias im Vergleich von  $n = 100$  zu  $n = 1000$  nun abnimmt (s. Abb. 14S und Tab. 23 im Anhang). Im Falle ungerundeter P-Werte wiesen die Schätzmethoden 7, 8 und 9 hingegen nur geringe Abweichungen auf.

#### **Approximative $\text{var}(\log(HR))$ -Schätzer und relative Gewichte der Methoden 1 bis 9**

Im Gegensatz zur Schätzung der  $\log(HR)$ -Werte, war nach Berechnung der  $\text{var}(\log(HR))$ -Werte keine Biasbestimmung möglich, da keine „wahre“ Varianz vorgegeben werden konnte. Die sechs untersuchten  $\text{var}(\log(HR))$ -Approximationen schätzten allesamt geringere gepoolte Varianzen als die Cox-Regression, wobei die Methoden 5 und 6 bzw. 8 und 9 (InVindir3 und InVindir4 bzw. InVindir6 und InVindir7, Formeln 13 und 14) die geringsten gepoolten Varianzen schätzten (s. Tab. 3). Diese betragen nur 55% (Methode 5 bzw. 8) bzw. 70% (Methode 6 bzw. 9) der jeweiligen Schätzer der Cox-Regression. Dementsprechend würden Studien, deren Varianzschätzer mithilfe der vorgenannten Methoden extrahiert bzw. approximiert werden, im Verhältnis deutlich höhere relative Gewichte erhalten. Diese Varianzunterschätzungen und folglich Übergewichtungen schlugen, wie von Parmar vermutet, insbesondere im Falle niedrigerer Hazard Ratios und ungleicher

Allokationsverhältnisse zu Buche (vgl. Abb. 16 „Allocation Ratios“ und „True Hazard Ratios“ sowie Tab. 24 und 27 im Anhang).<sup>9</sup> Methode 5 bzw. 8 (InVindir3 bzw. InVindir6, Formel 13) stach mit den niedrigsten Varianzschätzern unter Allokationsimbalance hervor (vgl. Abb. 16 „Allocation Ratios“ und Tab. 24 im Anhang). Anfällig für den Einfluss unbalanzierter Allokationsraten zeigten sich außerdem die Methoden 1 und 2 (InVdir1 und InVdir2, Formeln 9 und 10), die unter Imbalance zugunsten der Kontrollgruppe niedrige Varianzen und hohe Gewichte schätzten, sowie unter Imbalance zugunsten der Behandlungsgruppe hohe Varianzen und niedrige Gewichte schätzten (vgl. Abb. 16 „Allocation Ratios“ und Tab. 24 im Anhang). Die auf berichteten Konfidenzintervallen basierende Varianzschätzmethode 3 (InVindir1, Formel 11) schätzte dahingegen durchwegs sehr ähnliche Varianzen wie die Cox-Regression und wich in keinem der simulierten Szenarien mehr als 5% von diesen ab. Der größte relative Unterschied von etwas über 4% ergab sich bei einem vorgegebenem Hazard Ratio von 0.1 (vgl. Tab. 27 im Anhang). Der auf beobachteten Ereignissen beruhende Varianzschätzer der Methoden 4 und 7 (InVindir2, Formel 12) wich insbesondere in Szenarien mit großen Therapieeffekten und zusätzlichen Zensierungen von den Werten der Cox-Regression ab (s. Tab. 26 und 26 im Anhang) und zeigte ansonsten moderate Abweichungen. Die MCSE der Varianzschätzungen lagen allesamt unter 0.001.

### *Anwendungsempfehlungen für die Praxis*

In Anbetracht der vorgenannten beobachteten Ergebnisse der vorliegenden Simulationsstudie und daran angelehnten Diskussion der Performance der  $\log(HR)$ - und  $\text{var}(\log(HR))$ -Schätzmethoden lassen sich folgende Anwendungsempfehlung ableiten:

1. Generell nimmt die absolute mittlere Abweichung der  $\log(HR)$ -Schätzer (Bias) mit kleineren Therapieeffekten ab.
2. Ab einem HR von 0.5 sind nur noch geringe Abweichungen zu erwarten.
3. Die relative Abweichung betrug ab einem simulierten  $HR_{\text{true}}$  von 0.5 maximal 12%.
4. Die Basisereignisrate (Baseline Hazard) hat keinen relevanten Einfluss auf die Genauigkeit der  $\log(HR)$ - und  $\text{var}(\log(HR))$ -Schätzer.
5. Die direkte  $\log(HR)$ -Berechnungsmethode 1 (Formel 4) liefert die verlässlichsten Schätzer und ist zu präferieren.

6. Die direkte  $\log(HR)$ -Berechnungsmethode 2 (Formel 5) sollte nur bei balancierten Allokationsverhältnissen verwendet werden.
7. Die auf berichteten P-Werten basierenden  $\log(HR)$ -Schätzer 7 bis 9 sollten nur beim Vorhandensein exakter P-Werte verwendet werden (Formeln 6 bis 8).
8. Die  $\text{var}(\log(HR))$ -Approximation über berichtete Konfidenzintervalle (Formel 11) liegt sehr nahe an den Werten der Cox-Regression und ist ausnahmslos zu empfehlen.
9. Die direkten  $\text{var}(\log(HR))$ -Schätzmethoden 1 und 2 liegen nahe an den Werten der Cox-Regression, sollte jedoch nur bei balancierten Allokationsverhältnissen verwendet werden, wobei Methode 2 (Formel 10) Methode 1 (Formel 9) vorzuziehen ist.
10. Der indirekte Varianzschätzer der Methoden 4 und 7 (Formel 12) ist bedingt zu empfehlen und nur bei einem geringen Anteil von Zensierungen und nicht allzu starkem Therapieeffekt.
11. Die indirekten Varianzschätzer der Methoden 5 und 6 bzw. 8 und 9 unterschätzen  $\text{var}(\log(HR))$  stark (Formeln 13 und 14) und sind nicht zu empfehlen.

### *Limitationen der vorliegenden Simulationsstudie*

Trotz nicht von der Hand zu weisender Vorteile weisen Simulationsstudien im Allgemeinen auch spezifische Nachteile auf. Darüber hinaus schwächen einige methodische Besonderheiten die Generalisierbarkeit der vorliegenden Simulationsstudie im Speziellen. Beide Aspekte werden nun abschließend beleuchtet:

#### **Verwendung mittels Pseudozufallszahlen simulierter Datensätze**

Der größte Vorteil simulierter Datensätze ist zugleich auch der größte Nachteil von biostatistischen Simulationsstudien. Zunächst einmal bedeutet die Datensimulation mittels Pseudozufallszahlengeneratoren einen nicht zu unterschätzenden Rechenaufwand.<sup>23,24</sup> Im Falle der vorliegenden Simulationsstudie betrug die benötigte Rechenzeit über 12 Tage. Dies bedeutet in der schnelllebigen Wissenschaftsrealität mit ihrem immensen Zeitdruck eine wichtige Limitation bei der Verwendung von Simulationsstudien, da zur Minimierung der Rechenzeit ggf. auf Kosten der zu untersuchenden Parameterkombinationen oder zu Lasten der Komplexität der zu untersuchenden Methoden Abstriche gemacht werden. Lösungsansätze bietet hier das Feld des „Parallel Computing“, was aber nicht zuletzt im Falle von R-Code einen erheblichen sekundären Umformulierungsaufwand bedeuten kann.<sup>32,33</sup> Im Rahmen

der vorliegenden Simulationsstudie wurden die zu untersuchenden Parameterkombinationen allerdings a priori definiert und auch in Anbetracht der langen Rechenzeit nicht verändert. Neben der Rechenintensität bedeutet die Simulation von Daten darüber hinaus immer auch eine Simplifikation hochkomplexer biologischer Zusammenhänge, weshalb Simulationsstudien trotz der ehrenvollen Absicht der Biasuntersuchung damit selbst äußerst anfällig für ein nicht zu unterschätzendes Verzerrungspotential sind.<sup>23,24</sup> Obwohl durch die artifizielle Konstruktion von Parameterkombinationen eine Vielzahl an Szenarien systematisch untersucht werden kann, besteht zudem die Gefahr schlecht generalisierbare oder für die klinische Praxis vollkommen irrelevante Daten zu erzeugen.<sup>23,24</sup> Der vorliegenden Simulationsstudie diente allerdings eine klinisch höchst relevante Problemstellung als Grundlage und die Simulation der Survivaldaten wurde gemäß etablierter Standards inkl. eines Rekrutierungszeitraumes und verschiedener Zensierungsformen durchgeführt.<sup>10-12,22,34</sup> Nichtsdestotrotz kann nicht ausgeschlossen werden, dass die verwendeten Daten-generierende Modelle jedweder Simulationsstudie die wahren Gegebenheiten nicht adäquat widerspiegeln. Insbesondere wurden alle zur Berechnung der Schätzmethoden benötigten Größen simuliert, sodass diese uneingeschränkt verglichen werden konnten, was in der Realität unzureichend berichteter Überlebenszeitstudien nicht der Fall ist.

### **Verwendung exponentialverteilter Überlebenszeitdaten, proportionaler Hazards und exponential-/uniformverteilter Zensierungen**

Im Rahmen der vorliegenden Simulationsstudie wurden Überlebenszeitdaten univariat exponentialverteilt simuliert, um Annahmen der Cox-Regression auf der einen und der Log-Rang-Methode auf der anderen Seite nicht grob zu verletzen, wobei dieser Ansatz der gängigen Praxis in diesem Kontext entspricht.<sup>35,36</sup> Allerdings impliziert dieser Ansatz selbst eine sehr konservative Verteilungsannahme, die in der klinischen Realität nicht immer zutrifft, insbesondere im Falle zeitabhängiger Effekte und inkonstanter oder nicht-proportionaler Ausfallraten.<sup>36-38</sup> Dementsprechend muss die externe Validität der so gearteten Modellierung von Überlebenszeitdaten kritisch beurteilt werden bzw. ggf. Alternativen bemüht werden.<sup>39,40</sup> Da jedoch postuliert werden kann, dass die überwiegende Mehrzahl publizierter Überlebenszeitstudien auf der Log-Rang-Methode oder der Cox-Regression unter Annahme proportionaler Hazards beruht, ist die vorliegende Simulationsstudie folgerichtig auch in der

überwiegenden Mehrheit der Fälle etwaiger Metaanalysen in diesem Zusammenhang von Relevanz.<sup>2,12,41</sup> Neben den univariat exponentialverteilt generierten Überlebenszeiten müssen auch die simulierten Zensierungsmechanismen diskutiert werden. Diese entsprachen zum einen Rechtszensierungen bei Erreichen der Follow-Up Zeit ohne zusätzliche Zensierungen, sowie zum anderen zusätzliche exponential- und stetig gleichverteilte Zensierungen. Dieser Herangehensweise liegt das Bestreben zugrunde nicht-informative, unabhängige und zufällige Zensierungsmechanismen zu simulieren, was man in der klinischen Forschung mithilfe des Goldstandards einer doppelblinden randomisierten kontrollierten Studie zu erreichen versucht. Der gewählte Simulationsansatz entspricht dabei den einschlägigen Empfehlungen, kann aber die volle Bandbreite von in der Realität anzutreffenden Zensierungsformen nicht abbilden.<sup>22,42,43</sup>

#### *Ausblick und Ansatzpunkte für weiterführende Untersuchungen*

Auf Grundlage der vorgebrachten Limitationen der vorliegenden Simulationsstudie muss zum einen gefordert werden, dass die untersuchten Methoden auch in Szenarien mit anders gearteten Datenverteilungen und unter Verletzung wichtiger Annahmen getestet werden, bspw. unter nicht-proportionalen Hazards und inkonstanten Therapieeffekten. Zum anderen sollte die vorliegende Simulationsstudie und der ihr zugrundeliegende Code öffentlich gemacht werden, um eine breite Diskussion der Methoden und Ergebnisse mit maximaler Transparenz zu garantieren und nicht zuletzt eben diese weiterführenden Untersuchungen zu ermöglichen.

In diesem Sinne möchte ich ein auch von Tim Morris referenziertes Zitat hier anbringen: „*Simulation studies reveal points of light on a landscape, but cannot illuminate the entire landscape.*”<sup>24</sup>



## Abbildungsverzeichnis

Abb. 1: Zusammenhand zwischen $h(t)$ , $S(t)$ und $H(t)$ .....	S. 11
Abb. 2: Hazard Ratios quantifizieren die Verhältnisse von Ausfallraten .....	S. 11
Abb. 3: Log-Transformation zur linearen Darstellung von Hazard Ratios .....	S. 14
Abb. 4: Ridgeplots der gepoolten $\log(\text{HR})$ -Punktschätzer .....	S. 39
Abb. 5: Fehlerbalkendiagramme der Abweichungen der $\log(\text{HR})$ -Punktschätzer .....	S. 40
Abb. 6: Abweichungen der $\log(\text{HR})$ -Schätzer bei vorgegebenem Hazard Ratio von 0.1 -	S. 42
Abb. 7: Ridgeplots der $\log(\text{HR})$ -Schätzer bei vorgegebenem Hazard Ratio von 0.1 -----	S. 43
Abb. 8: Abweichungen der $\log(\text{HR})$ -Schätzer bei vorgegebenem Hazard Ratio von 0.3 -	S. 44
Abb. 9: Ridgeplots der $\log(\text{HR})$ -Schätzer bei vorgegebenem Hazard Ratio von 0.3 -----	S. 45
Abb. 10: Abweichungen der $\log(\text{HR})$ -Schätzer bei vorgegebenem Hazard Ratio von 0.5 -	S. 46
Abb. 11: Ridgeplots der $\log(\text{HR})$ -Schätzer bei vorgegebenem Hazard Ratio von 0.5 -----	S. 47
Abb. 12: Abweichungen der $\log(\text{HR})$ -Schätzer bei vorgegebenem Hazard Ratio von 0.7 -	S. 48
Abb. 13: Ridgeplots der $\log(\text{HR})$ -Schätzer bei vorgegebenem Hazard Ratio von 0.7 -----	S. 49
Abb. 14: Abweichungen der $\log(\text{HR})$ -Schätzer bei vorgegebenem Hazard Ratio von 0.9 -	S. 50
Abb. 15: Ridgeplots der $\log(\text{HR})$ -Schätzer bei vorgegebenem Hazard Ratio von 0.9 -----	S. 51
Abb. 16: Scatterplots simulierter relativer Studiengewichte .....	S. 53

## Tabellenverzeichnis

Tabelle 1: ADEMP-Struktur der vorliegenden Simulationsstudie -----	S. 18-19
Tabelle 2: Über alle Szenarien gemittelte Punktschätzer für $\log(HR)$ -----	S. 38
Tabelle 3: Über alle Szenarien gemittelte Punktschätzer für $\text{var}(\log(HR))$ -----	S. 52
Tabellen 4 bis 28: siehe Anhang -----	S. 73-99

## Literaturverzeichnis

1. Altman DG, Bland JM. Statistics Notes: Time to event (survival) data. *BMJ*. 1998;317(7156):468-469. doi:10.1136/bmj.317.7156.468
2. Clark TG, Bradburn MJ, Love SB, Altman DG. Survival Analysis Part I: Basic concepts and first analyses. *Br J Cancer*. 2003;89(2):232-238. doi:10.1038/sj.bjc.6601118
3. Guyatt G. Evidence-Based Medicine. *JAMA*. 1992;268(17):2420. doi:10.1001/jama.1992.03490170092032
4. Oxman AD, Guyatt GH. The Science of Reviewing Research. *Ann N Y Acad Sci*. 1993;703(1 Doing More Go):125-134. doi:10.1111/j.1749-6632.1993.tb26342.x
5. Egger M, Smith GD. Meta-analysis: Potentials and promise. *Bmj*. 1997;315(7119):1371-1374. doi:10.1136/bmj.315.7119.1371
6. Normand S-LT. Meta-analysis: formulating, evaluating, combining, and reporting. *Stat Med*. 1999;18(3):321-359. doi:10.1002/(SICI)1097-0258(19990215)18:3<321::AID-SIM28>3.0.CO;2-P
7. Ziegler A, Lange S, Bender R. Systematische Übersichten und Meta-Analysen. *DMW - Dtsch Medizinische Wochenschrift*. 2007;132(S 01):e48-e52. doi:10.1055/s-2007-959042
8. Borenstein M, Hedges L V., Higgins JPT, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Methods*. 2010;1(2):97-111. doi:10.1002/jrsm.12
9. Parmar MKB, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Stat Med*. 1998;17(24):2815-2834. doi:10.1002/(SICI)1097-0258(19981230)17:24<2815::AID-SIM110>3.0.CO;2-8
10. Mathoulin-Pelissier S, Gourgou-Bourgade S, Bonnetain F, Kramar A. Survival end point reporting in randomized cancer clinical trials: A review of major journals. *J Clin Oncol*. 2008;26(22):3721-3726. doi:10.1200/JCO.2007.14.1192
11. Abaira V, Muriel A, Emparanza JI, et al. Reporting quality of survival analyses in medical journals still needs improvement. A minimal requirements proposal. *J Clin Epidemiol*. 2013;66(12):1340-1346.e5. doi:10.1016/j.jclinepi.2013.06.009
12. Batson S, Greenall G, Hudson P. Review of the Reporting of Survival Analyses

- within Randomised Controlled Trials and the Implications for Meta-Analysis. Virgili G, ed. *PLoS One*. 2016;11(5):e0154870.  
doi:10.1371/journal.pone.0154870
13. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ*. 2009;339(jul21 1):b2700-b2700. doi:10.1136/bmj.b2700
  14. Hoffmann F, Allers K, Rombey T, et al. Nearly 80 systematic reviews were published each day: Observational study on trends in epidemiology and reporting over the years 2000-2019. *J Clin Epidemiol*. 2021;138:1-11. doi:10.1016/j.jclinepi.2021.05.022
  15. Kovalchik SA. Survey finds that most meta-analysts do not attempt to collect individual patient data. *J Clin Epidemiol*. 2012;65(12):1296-1299. doi:10.1016/j.jclinepi.2012.07.010
  16. Tudur C, Williamson PR, Khan S, Best LY. The value of the aggregate data approach in meta-analysis with time-to-event outcomes. *J R Stat Soc*. 2001;164(2):357-370. doi:10.1111/1467-985X.00207
  17. Williamson PR, Smith CT, Hutton JL, Marson AG. Aggregate data meta-analysis with time-to-event outcomes. *Stat Med*. 2002;21(22):3337-3351. doi:10.1002/sim.1303
  18. Michiels S, Piedbois P, Burdett S, Syz N, Stewart L, Pignon J-P. Meta-analysis when only the median survival times are known: A comparison with individual patient data results. *Int J Technol Assess Health Care*. 2005;21(1):119-125. doi:10.1017/S0266462305050154
  19. Tierney JF, Stewart LA, Ghersi D, Burdett S, Sydes MR. Practical methods for incorporating summary time-to-event data into meta-analysis. *Trials*. 2007;8(1):16. doi:10.1186/1745-6215-8-16
  20. Tierney JF, Fisher DJ, Burdett S, Stewart LA, Parmar MKB. Comparison of aggregate and individual participant data approaches to meta-analysis of randomised trials: An observational study. Shapiro SD, ed. *PLOS Med*. 2020;17(1):e1003019. doi:10.1371/journal.pmed.1003019
  21. Hebert AE, Kreaden US, Yankovsky A, et al. Methodology to standardize heterogeneous statistical data presentations for combining time-to-event oncologic outcomes. Pathak M, ed. *PLoS One*. 2022;17(2):e0263661.

- doi:10.1371/journal.pone.0263661
22. Burton A, Altman DG, Royston P, Holder RL. The design of simulation studies in medical statistics. *Stat Med.* 2006;25(24):4279-4292. doi:10.1002/sim.2673
  23. Boulesteix A-L, Groenwold RHH, Abrahamowicz M, et al. Introduction to statistical simulations in health research. *BMJ Open.* 2020;10(12):e039921. doi:10.1136/bmjopen-2020-039921
  24. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. *Stat Med.* 2019;38(11):2074-2102. doi:10.1002/sim.8086
  25. Cox DR. Regression Models and Life-Tables. *J R Stat Soc Ser B.* 1972;34(2):187-202. doi:10.1111/j.2517-6161.1972.tb00899.x
  26. American Psychological Association. APA Style numbers and statistics guide. <https://apastyle.apa.org/instructional-aids/numbers-statistics-guide.pdf>. Published 2022. Accessed February 2, 2023.
  27. Peto R, Pike MC, Armitage P, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. Analysis and examples. *Br J Cancer.* 1977;35(1):1-39. doi:10.1038/bjc.1977.1
  28. Lin DY, Dai L, Cheng G, Sailer MO. On confidence intervals for the hazard ratio in randomized clinical trials. *Biometrics.* 2016;72(4):1098-1102. doi:10.1111/biom.12528
  29. Joshi M, Pustejovsky James. simhelpers: Helper Functions for Simulation Studies. 2022. <https://cran.r-project.org/package=simhelpers>.
  30. Schoenfeld DA. Sample-Size Formula for the Proportional-Hazards Regression Model. *Biometrics.* 1983;39(2):499. doi:10.2307/2531021
  31. Persson I, Khamis H. Bias Of The Cox Model Hazard Ratio. *J Mod Appl Stat Methods.* 2005;4(1):90-99. doi:10.22237/jmasm/1114906200
  32. Ševčíková H. Statistical Simulations on Parallel Computers. *J Comput Graph Stat.* 2004;13(4):886-906. doi:10.1198/106186004X12605
  33. Melab N, Gmys J, Korosec P, Chakroun I. Synergy between parallel computing, optimization and simulation. *J Comput Sci.* 2020;44:101168. doi:10.1016/j.jocs.2020.101168
  34. Higgins JP, Li T, Deeks JJ. Choosing effect measures and computing estimates of effect. In: Higgins J, Thomas J, Chandler J, et al., eds. *Cochrane Handbook for Systematic Reviews of Interventions*. 2nd Editio. Chichester (UK): John Wiley & Sons; 2019. doi:10.1002/9781119536604

35. Bender R, Augustin T, Blettner M. Generating survival times to simulate Cox proportional hazards models. *Stat Med.* 2005;24(11):1713-1723. doi:10.1002/sim.2059
36. Austin PC. Generating survival times to simulate Cox proportional hazards models with time-varying covariates. *Stat Med.* 2012;31(29):3946-3958. doi:10.1002/sim.5452
37. Jachno K, Heritier S, Wolfe R. Impact of a non-constant baseline hazard on detection of time-dependent treatment effects: a simulation study. *BMC Med Res Methodol.* 2021;21(1):177. doi:10.1186/s12874-021-01372-0
38. Taketomi N, Yamamoto K, Chesneau C, Emura T. Parametric Distributions for Survival and Reliability Analyses, a Review and Historical Sketch. *Mathematics.* 2022;10(20):3907. doi:10.3390/math10203907
39. Bradburn MJ, Clark TG, Love SB, Altman DG. Survival Analysis Part III: Multivariate data analysis – choosing a model and assessing its adequacy and fit. *Br J Cancer.* 2003;89(4):605-611. doi:10.1038/sj.bjc.6601120
40. Clark TG, Bradburn MJ, Love SB, Altman DG. Survival analysis part IV: Further concepts and methods in survival analysis. *Br J Cancer.* 2003;89(5):781-786. doi:10.1038/sj.bjc.6601117
41. Bradburn MJ, Clark TG, Love SB, Altman DG. Survival Analysis Part II: Multivariate data analysis – an introduction to concepts and methods. *Br J Cancer.* 2003;89(3):431-436. doi:10.1038/sj.bjc.6601119
42. Wan F. Simulating survival data with predefined censoring rates for proportional hazards models. *Stat Med.* 2017;36(5):838-854. doi:10.1002/sim.7178
43. Turkson AJ, Ayiah-Mensah F, Nimoh V. Handling Censoring and Censored Data in Survival Analysis: A Standalone Systematic Literature Review. Tang N, ed. *Int J Math Math Sci.* 2021;2021:1-16. doi:10.1155/2021/9307475

## Verzeichnis verwendeter R-Pakete

Die Software „RStudio“ wurde zur Umsetzung des Statistikprogrammes „R“ der Version 4.1.3 (2022-03-10) mit den dazugehörigen vorinstallierten Basispaketen verwendet.

Zusätzlich wurden folgende Pakete geladen:

- tidyverse
- magrittr
- janitor
- survival
- survminer
- svMisc
- progress
- knitr
- kableExtra
- ggribges
- ggh4x
- plotrix
- viridis
- gghighlight
- rockchalk

## Anhang

- Tabelle 4: Bias der  $\log(HR)$ -Schätzer 1 bis 9 mit  $HR_{true} = 0.1$  unter Variation der Allokationsrate
- Tabelle 5: Bias der  $\log(HR)$ -Schätzer 1 bis 9 mit  $HR_{true} = 0.1$  unter Variation des Baselinehazard
- Tabelle 6: Bias der  $\log(HR)$ -Schätzer 1 bis 9 mit  $HR_{true} = 0.1$  unter Variation der Zensierungsrate
- Tabelle 7: Bias der  $\log(HR)$ -Schätzer 1 bis 9 mit  $HR_{true} = 0.1$  unter Variation der Gesamtfallzahl
- Tabelle 8: Bias der  $\log(HR)$ -Schätzer 1 bis 9 mit  $HR_{true} = 0.3$  unter Variation der Allokationsrate
- Tabelle 9: Bias der  $\log(HR)$ -Schätzer 1 bis 9 mit  $HR_{true} = 0.3$  unter Variation des Baselinehazard
- Tabelle 10: Bias der  $\log(HR)$ -Schätzer 1 bis 9 mit  $HR_{true} = 0.3$  unter Variation der Zensierungsrate
- Tabelle 11: Bias der  $\log(HR)$ -Schätzer 1 bis 9 mit  $HR_{true} = 0.3$  unter Variation der Gesamtfallzahl
- Tabelle 12: Bias der  $\log(HR)$ -Schätzer 1 bis 9 mit  $HR_{true} = 0.5$  unter Variation der Allokationsrate
- Tabelle 13: Bias der  $\log(HR)$ -Schätzer 1 bis 9 mit  $HR_{true} = 0.5$  unter Variation des Baselinehazard
- Tabelle 14: Bias der  $\log(HR)$ -Schätzer 1 bis 9 mit  $HR_{true} = 0.5$  unter Variation der Zensierungsrate
- Tabelle 15: Bias der  $\log(HR)$ -Schätzer 1 bis 9 mit  $HR_{true} = 0.5$  unter Variation der Gesamtfallzahl
- Tabelle 16: Bias der  $\log(HR)$ -Schätzer 1 bis 9 mit  $HR_{true} = 0.7$  unter Variation der Allokationsrate
- Tabelle 17: Bias der  $\log(HR)$ -Schätzer 1 bis 9 mit  $HR_{true} = 0.7$  unter Variation des Baselinehazard
- Tabelle 18: Bias der  $\log(HR)$ -Schätzer 1 bis 9 mit  $HR_{true} = 0.7$  unter Variation der Zensierungsrate
- Tabelle 19: Bias der  $\log(HR)$ -Schätzer 1 bis 9 mit  $HR_{true} = 0.7$  unter Variation der Gesamtfallzahl
- Tabelle 20: Bias der  $\log(HR)$ -Schätzer 1 bis 9 mit  $HR_{true} = 0.9$  unter Variation der Allokationsrate
- Tabelle 21: Bias der  $\log(HR)$ -Schätzer 1 bis 9 mit  $HR_{true} = 0.9$  unter Variation des Baselinehazard
- Tabelle 22: Bias der  $\log(HR)$ -Schätzer 1 bis 9 mit  $HR_{true} = 0.9$  unter Variation der Zensierungsrate
- Tabelle 23: Bias der  $\log(HR)$ -Schätzer 1 bis 9 mit  $HR_{true} = 0.9$  unter Variation der Gesamtfallzahl
- Tabelle 24: Werte der  $var(\log(HR))$ -Punktschätzer 1 bis 9 unter Variation der Allokationsrate
- Tabelle 25: Werte der  $var(\log(HR))$ -Punktschätzer 1 bis 9 unter Variation des Baselinehazard
- Tabelle 26: Werte der  $var(\log(HR))$ -Punktschätzer 1 bis 9 unter Variation der Zensierungsrate
- Tabelle 27: Werte der  $var(\log(HR))$ -Punktschätzer 1 bis 9 unter Variation des Hazard Ratio
- Tabelle 28: Werte der  $var(\log(HR))$ -Punktschätzer 1 bis 9 unter Variation der Gesamtfallzahl



– Anhang –

Bias for Methods 1 through 9 with HR true = 0.1 [log(HR\_true) = -2.303]

Method	Allocation Ratio	Iterations	Theta	Variance	Bias	MCSE	Lower	Upper	MSE	RMSE
<b>COX PH</b>										
InHRobs	4:1	26284	-2.41513	0.16620	-0.11213	0.00251	-0.11706	-0.10720	0.17877	0.42281
InHRobs	2:1	26680	-2.41368	0.13243	-0.11068	0.00223	-0.11505	-0.10632	0.14467	0.38036
InHRobs	1:1	26876	-2.41165	0.11175	-0.10865	0.00204	-0.11265	-0.10466	0.12355	0.35150
InHRobs	1:2	26952	-2.41039	0.10650	-0.10739	0.00199	-0.11129	-0.10350	0.11802	0.34355
InHRobs	1:4	26955	-2.40379	0.11679	-0.10079	0.00208	-0.10487	-0.09671	0.12694	0.35629
<b>Method 1</b>										
InHRdir1	4:1	26284	-1.64937	0.20366	0.65363	0.00278	0.64818	0.65909	0.63089	0.79428
InHRdir1	2:1	26680	-1.67109	0.17828	0.63191	0.00258	0.62685	0.63698	0.57758	0.75999
InHRdir1	1:1	26876	-1.73670	0.14153	0.56630	0.00229	0.56181	0.57080	0.46222	0.67987
InHRdir1	1:2	26952	-1.84605	0.09820	0.45695	0.00191	0.45321	0.46069	0.30700	0.55408
InHRdir1	1:4	26955	-1.97509	0.07312	0.32791	0.00165	0.32468	0.33114	0.18064	0.42501
<b>Method 2</b>										
InHRdir2	4:1	26284	-1.50942	0.05577	0.79358	0.00146	0.79072	0.79643	0.68553	0.82797
InHRdir2	2:1	26680	-1.85511	0.09553	0.44789	0.00189	0.44418	0.45159	0.29613	0.54418
InHRdir2	1:1	26876	-2.43846	0.18431	-0.13546	0.00262	-0.14059	-0.13033	0.20265	0.45017
InHRdir2	1:2	26952	-3.36320	0.41293	-1.06020	0.00391	-1.06787	-1.05253	1.53694	1.23973
InHRdir2	1:4	26955	-4.62614	1.03289	-2.32314	0.00619	-2.33527	-2.31100	6.42982	2.53571
<b>Method 7</b>										
InHRindir5	4:1	26284	-0.99448	0.48462	1.30852	0.00429	1.30010	1.31693	2.19681	1.48217
InHRindir5	2:1	26680	-0.86276	0.42589	1.44024	0.00400	1.43241	1.44807	2.50017	1.58119
InHRindir5	1:1	26876	-0.79165	0.36418	1.51135	0.00368	1.50413	1.51856	2.64835	1.62737
InHRindir5	1:2	26952	-0.78812	0.31200	1.51488	0.00340	1.50821	1.52155	2.60685	1.61457
InHRindir5	1:4	26955	-0.86010	0.30646	1.44290	0.00337	1.43629	1.44950	2.38839	1.54544
<b>Method 8</b>										
InHRindir6	4:1	26284	-0.57720	0.08460	1.72580	0.00179	1.72228	1.72931	3.06297	1.75014
InHRindir6	2:1	26680	-0.61538	0.10825	1.68762	0.00201	1.68367	1.69156	2.95630	1.71939
InHRindir6	1:1	26876	-0.65807	0.14383	1.64493	0.00231	1.64040	1.64946	2.84962	1.68808
InHRindir6	1:2	26952	-0.70694	0.19512	1.59606	0.00269	1.59079	1.60134	2.74252	1.65606
InHRindir6	1:4	26955	-0.75320	0.25213	1.54980	0.00306	1.54381	1.55580	2.65401	1.62911
<b>Method 9</b>										
InHRindir7	4:1	26284	-0.72150	0.13219	1.58150	0.00224	1.57710	1.58589	2.63331	1.62275
InHRindir7	2:1	26680	-0.65356	0.12267	1.64944	0.00214	1.64524	1.65364	2.84331	1.68621
InHRindir7	1:1	26876	-0.65807	0.14383	1.64493	0.00231	1.64040	1.64946	2.84962	1.68808
InHRindir7	1:2	26952	-0.75083	0.22093	1.55217	0.00286	1.54656	1.55778	2.63016	1.62178
InHRindir7	1:4	26955	-0.94150	0.39395	1.36150	0.00382	1.35401	1.36900	2.24763	1.49921

– Anhang –

Bias for Methods 1 through 9 with HR true = 0.1 [ $\log(\text{HR}_{\text{true}}) = -2.303$ ]

Method	Baseline		Theta	Variance	Bias	MCSE	Lower	Upper	MSE	RMSE
	Hazard	Iterations								
<b>COX PH</b>										
InHRobs	25%	44148	-2.40473	0.14169	-0.10173	0.00179	-0.10524	-0.09822	0.15204	0.38992
InHRobs	50%	44712	-2.41265	0.12106	-0.10965	0.00165	-0.11288	-0.10643	0.13308	0.36480
InHRobs	75%	44887	-2.41523	0.11706	-0.11223	0.00161	-0.11539	-0.10906	0.12965	0.36007
<b>Method 1</b>										
InHRdir1	25%	44148	-1.81426	0.16956	0.48874	0.00196	0.48490	0.49258	0.40843	0.63908
InHRdir1	50%	44712	-1.77160	0.15526	0.53140	0.00186	0.52774	0.53505	0.43764	0.66154
InHRdir1	75%	44887	-1.74432	0.13254	0.55868	0.00172	0.55531	0.56205	0.44466	0.66683
<b>Method 2</b>										
InHRdir2	25%	44148	-2.72921	1.64610	-0.42621	0.00611	-0.43817	-0.41424	1.82772	1.35193
InHRdir2	50%	44712	-2.76959	1.59997	-0.46659	0.00598	-0.47831	-0.45486	1.81763	1.34820
InHRdir2	75%	44887	-2.80090	1.62673	-0.49790	0.00602	-0.50970	-0.48610	1.87459	1.36916
<b>Method 7</b>										
InHRindir5	25%	44148	-0.94080	0.48148	1.36220	0.00330	1.35573	1.36867	2.33706	1.52874
InHRindir5	50%	44712	-0.84647	0.36735	1.45653	0.00287	1.45091	1.46215	2.48883	1.57760
InHRindir5	75%	44887	-0.79037	0.29189	1.51263	0.00255	1.50763	1.51762	2.57992	1.60621
<b>Method 8</b>										
InHRindir6	25%	44148	-0.70538	0.19767	1.59762	0.00212	1.59347	1.60177	2.75006	1.65833
InHRindir6	50%	44712	-0.65539	0.15354	1.64761	0.00185	1.64398	1.65125	2.86817	1.69357
InHRindir6	75%	44887	-0.62795	0.12986	1.67505	0.00170	1.67171	1.67838	2.93563	1.71337
<b>Method 9</b>										
InHRindir7	25%	44148	-0.79325	0.26319	1.50975	0.00244	1.50497	1.51454	2.54254	1.59454
InHRindir7	50%	44712	-0.73716	0.20379	1.56584	0.00213	1.56166	1.57003	2.65566	1.62962
InHRindir7	75%	44887	-0.70670	0.17300	1.59630	0.00196	1.59245	1.60014	2.72116	1.64959

– Anhang –

Bias for Methods 1 through 9 with HR true = 0.1 [ $\log(\text{HR}_{\text{true}}) = -2.303$ ]

Method	Censoring Distribution	Iterations	Theta	Variance	Bias	MCSE	Lower	Upper	MSE	RMSE
<b>COX PH</b>										
InHRobs	none	44999	-2.32370	0.05595	-0.02070	0.00112	-0.02288	-0.01851	0.05638	0.23744
InHRobs	exp0.5	43854	-2.32477	0.15960	-0.02177	0.00191	-0.02551	-0.01803	0.16007	0.40009
InHRobs	unif0.5	44894	-2.58245	0.12072	-0.27945	0.00164	-0.28267	-0.27624	0.19881	0.44588
<b>Method 1</b>										
InHRdir1	none	44999	-1.52758	0.03632	0.77542	0.00090	0.77366	0.77718	0.63760	0.79850
InHRdir1	exp0.5	43854	-2.18013	0.12964	0.12287	0.00172	0.11950	0.12624	0.14474	0.38044
InHRdir1	unif0.5	44894	-1.63181	0.05112	0.67119	0.00107	0.66910	0.67328	0.50161	0.70824
<b>Method 2</b>										
InHRdir2	none	44999	-2.75799	1.26861	-0.45499	0.00531	-0.46540	-0.44458	1.47559	1.21474
InHRdir2	exp0.5	43854	-2.28711	1.06101	0.01589	0.00492	0.00625	0.02553	1.06124	1.03017
InHRdir2	unif0.5	44894	-3.24410	2.08058	-0.94110	0.00681	-0.95445	-0.92776	2.96621	1.72227
<b>Method 7</b>										
InHRindir5	none	44999	-0.53313	0.07639	1.76987	0.00130	1.76731	1.77242	3.20882	1.79132
InHRindir5	exp0.5	43854	-1.32293	0.59792	0.98007	0.00369	0.97283	0.98731	1.55844	1.24838
InHRindir5	unif0.5	44894	-0.73179	0.14912	1.57121	0.00182	1.56763	1.57478	2.61781	1.61796
<b>Method 8</b>										
InHRindir6	none	44999	-0.47212	0.05709	1.83088	0.00113	1.82868	1.83309	3.40922	1.84641
InHRindir6	exp0.5	43854	-0.87457	0.24001	1.42843	0.00234	1.42384	1.43301	2.28041	1.51010
InHRindir6	unif0.5	44894	-0.64671	0.10801	1.65629	0.00155	1.65325	1.65933	2.85130	1.68858
<b>Method 9</b>										
InHRindir7	none	44999	-0.53107	0.07549	1.77193	0.00130	1.76939	1.77447	3.21522	1.79310
InHRindir7	exp0.5	43854	-0.98414	0.32617	1.31886	0.00273	1.31352	1.32421	2.06555	1.43720
InHRindir7	unif0.5	44894	-0.72717	0.14225	1.57583	0.00178	1.57234	1.57932	2.62549	1.62034

– Anhang –

Bias for Methods 1 through 9 with HR true = 0.1 [ $\log(\text{HR}_{\text{true}}) = -2.303$ ]

Method	Trial Size	Iterations	Theta	Variance	Bias	MCSE	Lower	Upper	MSE	RMSE
<b>COX PH</b>										
InHRobs	N=100	43754	-2.43685	0.27421	-0.13385	0.00250	-0.13876	-0.12894	0.29212	0.54048
InHRobs	N=500	44993	-2.40328	0.06888	-0.10028	0.00124	-0.10271	-0.09786	0.07894	0.28096
InHRobs	N=1000	45000	-2.39329	0.03960	-0.09029	0.00094	-0.09213	-0.08845	0.04776	0.21853
<b>Method 1</b>										
InHRdir1	N=100	43754	-1.75169	0.19648	0.55131	0.00212	0.54715	0.55546	0.50042	0.70740
InHRdir1	N=500	44993	-1.78966	0.14153	0.51334	0.00177	0.50986	0.51681	0.40504	0.63643
InHRdir1	N=1000	45000	-1.78754	0.12183	0.51546	0.00165	0.51223	0.51868	0.38753	0.62252
<b>Method 2</b>										
InHRdir2	N=100	43754	-2.76856	1.86032	-0.46556	0.00652	-0.47834	-0.45278	2.07703	1.44119
InHRdir2	N=500	44993	-2.76383	1.53165	-0.46083	0.00583	-0.47226	-0.44939	1.74398	1.32060
InHRdir2	N=1000	45000	-2.76796	1.48962	-0.46496	0.00575	-0.47624	-0.45368	1.70578	1.30605
<b>Method 7</b>										
InHRindir5	N=100	43754	-1.43301	0.45919	0.86999	0.00324	0.86364	0.87634	1.21605	1.10275
InHRindir5	N=500	44993	-0.68141	0.13612	1.62159	0.00174	1.61818	1.62500	2.76566	1.66303
InHRindir5	N=1000	45000	-0.47779	0.06014	1.82521	0.00116	1.82295	1.82748	3.39153	1.84161
<b>Method 8</b>										
InHRindir6	N=100	43754	-1.11870	0.12878	1.18430	0.00172	1.18094	1.18766	1.53134	1.23747
InHRindir6	N=500	44993	-0.51669	0.02799	1.78631	0.00079	1.78476	1.78785	3.21888	1.79412
InHRindir6	N=1000	45000	-0.36525	0.01389	1.93775	0.00056	1.93666	1.93883	3.76875	1.94133
<b>Method 9</b>										
InHRindir7	N=100	43754	-1.25777	0.18687	1.04523	0.00207	1.04118	1.04928	1.27937	1.13109
InHRindir7	N=500	44993	-0.58153	0.04114	1.72147	0.00096	1.71960	1.72335	3.00461	1.73338
InHRindir7	N=1000	45000	-0.41122	0.02046	1.89178	0.00067	1.89046	1.89311	3.59931	1.89718

– Anhang –

Bias for Methods 1 through 9 with HR true = 0.3 [log(HR\_true) = -1.204]

Method	Allocation Ratio	Iterations	Theta	Variance	Bias	MCSE	Lower	Upper	MSE	RMSE
<b>COX PH</b>										
InHRobs	4:1	26923	-1.26643	0.09300	-0.06243	0.00186	-0.06608	-0.05879	0.09690	0.31129
InHRobs	2:1	26994	-1.26311	0.06681	-0.05911	0.00157	-0.06220	-0.05603	0.07030	0.26514
InHRobs	1:1	26999	-1.25948	0.05399	-0.05548	0.00141	-0.05825	-0.05270	0.05707	0.23888
InHRobs	1:2	27000	-1.26013	0.05507	-0.05613	0.00143	-0.05893	-0.05333	0.05822	0.24129
InHRobs	1:4	27000	-1.25949	0.06856	-0.05549	0.00159	-0.05862	-0.05237	0.07164	0.26766
<b>Method 1</b>										
InHRdir1	4:1	26923	-1.05039	0.06221	0.15361	0.00152	0.15063	0.15659	0.08580	0.29292
InHRdir1	2:1	26994	-1.05101	0.04418	0.15299	0.00128	0.15048	0.15550	0.06758	0.25996
InHRdir1	1:1	26999	-1.07150	0.03488	0.13250	0.00114	0.13028	0.13473	0.05244	0.22900
InHRdir1	1:2	27000	-1.11028	0.03441	0.09372	0.00113	0.09151	0.09594	0.04319	0.20782
InHRdir1	1:4	27000	-1.15194	0.04598	0.05206	0.00130	0.04950	0.05462	0.04869	0.22066
<b>Method 2</b>										
InHRdir2	4:1	26923	-0.97038	0.03494	0.23362	0.00114	0.23139	0.23586	0.08952	0.29920
InHRdir2	2:1	26994	-1.12146	0.04385	0.08254	0.00127	0.08005	0.08504	0.05066	0.22509
InHRdir2	1:1	26999	-1.33956	0.06875	-0.13556	0.00160	-0.13869	-0.13243	0.08712	0.29517
InHRdir2	1:2	27000	-1.62160	0.13604	-0.41760	0.00224	-0.42200	-0.41321	0.31043	0.55716
InHRdir2	1:4	27000	-1.91903	0.29559	-0.71503	0.00331	-0.72152	-0.70855	0.80685	0.89825
<b>Method 7</b>										
InHRindir5	4:1	26923	-0.75272	0.17688	0.45128	0.00256	0.44625	0.45630	0.38052	0.61686
InHRindir5	2:1	26994	-0.66449	0.14584	0.53951	0.00232	0.53495	0.54406	0.43691	0.66099
InHRindir5	1:1	26999	-0.62196	0.12168	0.58204	0.00212	0.57788	0.58620	0.46045	0.67857
InHRindir5	1:2	27000	-0.63819	0.11710	0.56581	0.00208	0.56172	0.56989	0.43724	0.66124
InHRindir5	1:4	27000	-0.70738	0.12585	0.49662	0.00216	0.49239	0.50085	0.37248	0.61032
<b>Method 8</b>										
InHRindir6	4:1	26923	-0.52667	0.05725	0.67733	0.00146	0.67447	0.68019	0.51602	0.71835
InHRindir6	2:1	26994	-0.56874	0.07786	0.63526	0.00170	0.63193	0.63859	0.48142	0.69384
InHRindir6	1:1	26999	-0.59652	0.09527	0.60748	0.00188	0.60380	0.61116	0.46430	0.68139
InHRindir6	1:2	27000	-0.61199	0.10723	0.59201	0.00199	0.58811	0.59592	0.45771	0.67654
InHRindir6	1:4	27000	-0.60727	0.10704	0.59673	0.00199	0.59282	0.60063	0.46312	0.68053
<b>Method 9</b>										
InHRindir7	4:1	26923	-0.65834	0.08946	0.54566	0.00182	0.54209	0.54923	0.38720	0.62225
InHRindir7	2:1	26994	-0.60400	0.08825	0.60000	0.00181	0.59646	0.60355	0.44825	0.66952
InHRindir7	1:1	26999	-0.59652	0.09527	0.60748	0.00188	0.60380	0.61116	0.46430	0.68139
InHRindir7	1:2	27000	-0.64994	0.12148	0.55406	0.00212	0.54990	0.55822	0.42846	0.65457
InHRindir7	1:4	27000	-0.75909	0.16725	0.44491	0.00249	0.44003	0.44978	0.36518	0.60430

– Anhang –

Bias for Methods 1 through 9 with HR true = 0.3 [ $\log(\text{HR}_{\text{true}}) = -1.204$ ]

Method	Baseline		Theta	Variance	Bias	MCSE	Lower	Upper	MSE	RMSE
	Hazard	Iterations								
<b>COX PH</b>										
InHRobs	25%	44928	-1.26435	0.08106	-0.06035	0.00134	-0.06298	-0.05771	0.08470	0.29103
InHRobs	50%	44991	-1.26076	0.06322	-0.05676	0.00119	-0.05909	-0.05444	0.06644	0.25776
InHRobs	75%	44997	-1.26008	0.05817	-0.05608	0.00114	-0.05831	-0.05385	0.06132	0.24762
<b>Method 1</b>										
InHRdir1	25%	44928	-1.09775	0.06064	0.10625	0.00116	0.10397	0.10853	0.07192	0.26818
InHRdir1	50%	44991	-1.08537	0.04169	0.11863	0.00096	0.11674	0.12052	0.05576	0.23614
InHRdir1	75%	44997	-1.07803	0.03504	0.12597	0.00088	0.12424	0.12770	0.05091	0.22563
<b>Method 2</b>										
InHRdir2	25%	44928	-1.38522	0.24810	-0.18122	0.00235	-0.18582	-0.17661	0.28094	0.53003
InHRdir2	50%	44991	-1.39571	0.22759	-0.19171	0.00225	-0.19612	-0.18730	0.26433	0.51413
InHRdir2	75%	44997	-1.40304	0.22223	-0.19904	0.00222	-0.20340	-0.19469	0.26184	0.51170
<b>Method 7</b>										
InHRindir5	25%	44928	-0.71737	0.16719	0.48663	0.00193	0.48285	0.49041	0.40400	0.63561
InHRindir5	50%	44991	-0.66774	0.13088	0.53626	0.00171	0.53292	0.53961	0.41845	0.64688
InHRindir5	75%	44997	-0.64568	0.11842	0.55832	0.00162	0.55514	0.56150	0.43015	0.65585
<b>Method 8</b>										
InHRindir6	25%	44928	-0.60808	0.09973	0.59592	0.00149	0.59300	0.59884	0.45484	0.67442
InHRindir6	50%	44991	-0.57669	0.08675	0.62731	0.00139	0.62459	0.63003	0.48027	0.69302
InHRindir6	75%	44997	-0.56208	0.08226	0.64192	0.00135	0.63927	0.64457	0.49432	0.70308
<b>Method 9</b>										
InHRindir7	25%	44928	-0.68246	0.12885	0.52154	0.00169	0.51822	0.52485	0.40085	0.63313
InHRindir7	50%	44991	-0.64738	0.11153	0.55662	0.00157	0.55354	0.55971	0.42135	0.64912
InHRindir7	75%	44997	-0.63093	0.10545	0.57307	0.00153	0.57007	0.57607	0.43385	0.65868

– Anhang –

Bias for Methods 1 through 9 with HR true = 0.3 [ $\log(\text{HR\_true}) = -1.204$ ]

Method	Censoring Distribution	Iterations	Theta	Variance	Bias	MCSE	Lower	Upper	MSE	RMSE
<b>COX PH</b>										
InHRobs	none	45000	-1.21199	0.03042	-0.00799	0.00082	-0.00960	-0.00638	0.03049	0.17461
InHRobs	exp0.5	44917	-1.22011	0.09399	-0.01611	0.00145	-0.01894	-0.01327	0.09425	0.30700
InHRobs	unif0.5	44999	-1.35301	0.06554	-0.14901	0.00121	-0.15137	-0.14664	0.08774	0.29621
<b>Method 1</b>										
InHRdir1	none	45000	-0.99622	0.01494	0.20778	0.00058	0.20665	0.20891	0.05811	0.24107
InHRdir1	exp0.5	44917	-1.16381	0.07795	0.04019	0.00132	0.03760	0.04277	0.07956	0.28207
InHRdir1	unif0.5	44999	-1.10125	0.03037	0.10275	0.00082	0.10114	0.10436	0.04093	0.20232
<b>Method 2</b>										
InHRdir2	none	45000	-1.36766	0.15676	-0.16366	0.00187	-0.16732	-0.16001	0.18354	0.42842
InHRdir2	exp0.5	44917	-1.24668	0.20571	-0.04268	0.00214	-0.04688	-0.03849	0.20752	0.45555
InHRdir2	unif0.5	44999	-1.56936	0.28243	-0.36536	0.00251	-0.37027	-0.36045	0.41591	0.64491
<b>Method 7</b>										
InHRindir5	none	45000	-0.50999	0.06730	0.69401	0.00122	0.69162	0.69641	0.54895	0.74091
InHRindir5	exp0.5	44917	-0.83226	0.18521	0.37174	0.00203	0.36776	0.37572	0.32340	0.56868
InHRindir5	unif0.5	44999	-0.68876	0.11463	0.51524	0.00160	0.51211	0.51836	0.38009	0.61652
<b>Method 8</b>										
InHRindir6	none	45000	-0.45371	0.05145	0.75029	0.00107	0.74820	0.75239	0.61438	0.78383
InHRindir6	exp0.5	44917	-0.68279	0.10417	0.52121	0.00152	0.51823	0.52420	0.37583	0.61305
InHRindir6	unif0.5	44999	-0.61050	0.08683	0.59350	0.00139	0.59078	0.59622	0.43907	0.66262
<b>Method 9</b>										
InHRindir7	none	45000	-0.50998	0.06730	0.69402	0.00122	0.69163	0.69642	0.54896	0.74092
InHRindir7	exp0.5	44917	-0.76587	0.13485	0.43813	0.00173	0.43473	0.44153	0.32680	0.57167
InHRindir7	unif0.5	44999	-0.68509	0.11089	0.51891	0.00157	0.51583	0.52198	0.38015	0.61656

– Anhang –

Bias for Methods 1 through 9 with HR true = 0.3 [ $\log(\text{HR}_{\text{true}}) = -1.204$ ]

Method	Trial Size	Iterations	Theta	Variance	Bias	MCSE	Lower	Upper	MSE	RMSE
<b>COX PH</b>										
InHRobs	N=100	44916	-1.27866	0.15370	-0.07466	0.00185	-0.07828	-0.07103	0.15927	0.39908
InHRobs	N=500	45000	-1.25453	0.03098	-0.05053	0.00083	-0.05216	-0.04890	0.03353	0.18313
InHRobs	N=1000	45000	-1.25203	0.01749	-0.04803	0.00062	-0.04925	-0.04681	0.01980	0.14070
<b>Method 1</b>										
InHRdir1	N=100	44916	-1.08046	0.09754	0.12354	0.00147	0.12065	0.12642	0.11280	0.33586
InHRdir1	N=500	45000	-1.08945	0.02432	0.11455	0.00074	0.11310	0.11599	0.03744	0.19349
InHRdir1	N=1000	45000	-1.09120	0.01571	0.11280	0.00059	0.11164	0.11395	0.02843	0.16862
<b>Method 2</b>										
InHRdir2	N=100	44916	-1.40398	0.37745	-0.19998	0.00290	-0.20567	-0.19430	0.41743	0.64609
InHRdir2	N=500	45000	-1.38990	0.17092	-0.18590	0.00195	-0.18972	-0.18208	0.20548	0.45330
InHRdir2	N=1000	45000	-1.39011	0.14982	-0.18611	0.00182	-0.18969	-0.18253	0.18446	0.42948
<b>Method 7</b>										
InHRindir5	N=100	44916	-1.07304	0.12439	0.13096	0.00166	0.12769	0.13422	0.14154	0.37621
InHRindir5	N=500	45000	-0.56156	0.03126	0.64244	0.00083	0.64081	0.64408	0.44399	0.66632
InHRindir5	N=1000	45000	-0.39686	0.01511	0.80714	0.00058	0.80601	0.80828	0.66658	0.81645
<b>Method 8</b>										
InHRindir6	N=100	44916	-0.92126	0.06479	0.28274	0.00120	0.28038	0.28509	0.14473	0.38043
InHRindir6	N=500	45000	-0.48379	0.01537	0.72021	0.00058	0.71906	0.72135	0.53407	0.73080
InHRindir6	N=1000	45000	-0.34239	0.00767	0.86161	0.00041	0.86080	0.86242	0.75005	0.86605
<b>Method 9</b>										
InHRindir7	N=100	44916	-1.03242	0.08585	0.17158	0.00138	0.16887	0.17429	0.11529	0.33954
InHRindir7	N=500	45000	-0.54392	0.02258	0.66008	0.00071	0.65869	0.66147	0.45829	0.67697
InHRindir7	N=1000	45000	-0.38510	0.01133	0.81890	0.00050	0.81792	0.81989	0.68193	0.82579



– Anhang –

Bias for Methods 1 through 9 with HR true = 0.5 [log(HR\_true) = -0.693]

Method	Allocation Ratio	Iterations	Theta	Variance	Bias	MCSE	Lower	Upper	MSE	RMSE
<b>COX PH</b>										
InHRobs	4:1	26983	-0.73156	0.07222	-0.03856	0.00164	-0.04177	-0.03536	0.07370	0.27149
InHRobs	2:1	26999	-0.72650	0.04901	-0.03350	0.00135	-0.03614	-0.03086	0.05013	0.22390
InHRobs	1:1	27000	-0.72380	0.04129	-0.03080	0.00124	-0.03322	-0.02838	0.04224	0.20551
InHRobs	1:2	27000	-0.72736	0.04416	-0.03436	0.00128	-0.03687	-0.03186	0.04534	0.21293
InHRobs	1:4	27000	-0.72630	0.05924	-0.03330	0.00148	-0.03620	-0.03040	0.06035	0.24566
<b>Method 1</b>										
InHRdir1	4:1	26983	-0.66560	0.05273	0.02740	0.00140	0.02466	0.03013	0.05348	0.23125
InHRdir1	2:1	26999	-0.66129	0.03449	0.03171	0.00113	0.02949	0.03392	0.03550	0.18840
InHRdir1	1:1	27000	-0.66591	0.03034	0.02709	0.00106	0.02501	0.02917	0.03107	0.17627
InHRdir1	1:2	27000	-0.68041	0.03352	0.01259	0.00111	0.01040	0.01477	0.03368	0.18351
InHRdir1	1:4	27000	-0.69190	0.04810	0.00110	0.00133	-0.00152	0.00371	0.04810	0.21931
<b>Method 2</b>										
InHRdir2	4:1	26983	-0.60828	0.03580	0.08472	0.00115	0.08246	0.08698	0.04298	0.20731
InHRdir2	2:1	26999	-0.66984	0.03503	0.02316	0.00114	0.02092	0.02539	0.03556	0.18859
InHRdir2	1:1	27000	-0.75156	0.04612	-0.05856	0.00131	-0.06112	-0.05600	0.04954	0.22259
InHRdir2	1:2	27000	-0.85204	0.07776	-0.15904	0.00170	-0.16236	-0.15571	0.10305	0.32101
InHRdir2	1:4	27000	-0.94562	0.15064	-0.25262	0.00236	-0.25725	-0.24799	0.21445	0.46309
<b>Method 7</b>										
InHRindir5	4:1	26983	-0.57257	0.07694	0.12043	0.00169	0.11712	0.12374	0.09145	0.30240
InHRindir5	2:1	26999	-0.51924	0.05856	0.17376	0.00147	0.17088	0.17665	0.08876	0.29792
InHRindir5	1:1	27000	-0.49810	0.05312	0.19490	0.00140	0.19216	0.19765	0.09110	0.30183
InHRindir5	1:2	27000	-0.51353	0.05289	0.17947	0.00140	0.17673	0.18222	0.08510	0.29171
InHRindir5	1:4	27000	-0.55922	0.05817	0.13378	0.00147	0.13090	0.13665	0.07606	0.27580
<b>Method 8</b>										
InHRindir6	4:1	26983	-0.42584	0.02893	0.26716	0.00104	0.26513	0.26919	0.10031	0.31671
InHRindir6	2:1	26999	-0.46789	0.03894	0.22511	0.00120	0.22276	0.22747	0.08962	0.29936
InHRindir6	1:1	27000	-0.49111	0.04741	0.20189	0.00133	0.19929	0.20449	0.08817	0.29693
InHRindir6	1:2	27000	-0.49225	0.05039	0.20075	0.00137	0.19807	0.20343	0.09069	0.30115
InHRindir6	1:4	27000	-0.46749	0.04692	0.22551	0.00132	0.22292	0.22809	0.09777	0.31268
<b>Method 9</b>										
InHRindir7	4:1	26983	-0.53230	0.04521	0.16070	0.00129	0.15816	0.16324	0.07103	0.26652
InHRindir7	2:1	26999	-0.49678	0.04410	0.19622	0.00128	0.19372	0.19873	0.08261	0.28741
InHRindir7	1:1	27000	-0.49111	0.04741	0.20189	0.00133	0.19929	0.20449	0.08817	0.29693
InHRindir7	1:2	27000	-0.52266	0.05706	0.17034	0.00145	0.16749	0.17318	0.08607	0.29338
InHRindir7	1:4	27000	-0.58437	0.07331	0.10863	0.00165	0.10541	0.11186	0.08511	0.29173

– Anhang –

Bias for Methods 1 through 9 with HR true = 0.5 [ $\log(\text{HR}_{\text{true}}) = -0.693$ ]

Method	Baseline Hazard	Iterations	Theta	Variance	Bias	MCSE	Lower	Upper	MSE	RMSE
<b>COX PH</b>										
InHRobs	25%	44982	-0.72776	0.06434	-0.03476	0.00120	-0.03711	-0.03242	0.06555	0.25603
InHRobs	50%	45000	-0.72834	0.04977	-0.03534	0.00105	-0.03740	-0.03328	0.05101	0.22586
InHRobs	75%	45000	-0.72522	0.04545	-0.03222	0.00100	-0.03419	-0.03025	0.04649	0.21561
<b>Method 1</b>										
InHRdir1	25%	44982	-0.67561	0.05074	0.01739	0.00106	0.01531	0.01948	0.05104	0.22593
InHRdir1	50%	45000	-0.67399	0.03669	0.01901	0.00090	0.01724	0.02078	0.03705	0.19250
InHRdir1	75%	45000	-0.66949	0.03244	0.02351	0.00085	0.02185	0.02518	0.03299	0.18163
<b>Method 2</b>										
InHRdir2	25%	44982	-0.76192	0.09261	-0.06892	0.00143	-0.07173	-0.06611	0.09736	0.31202
InHRdir2	50%	45000	-0.76785	0.08091	-0.07485	0.00134	-0.07748	-0.07222	0.08651	0.29413
InHRdir2	75%	45000	-0.76669	0.07808	-0.07369	0.00132	-0.07627	-0.07111	0.08350	0.28897
<b>Method 7</b>										
InHRindir5	25%	44982	-0.55253	0.07257	0.14047	0.00127	0.13798	0.14296	0.09230	0.30382
InHRindir5	50%	45000	-0.52973	0.05733	0.16327	0.00113	0.16106	0.16548	0.08398	0.28980
InHRindir5	75%	45000	-0.51532	0.05162	0.17768	0.00107	0.17558	0.17978	0.08319	0.28843
<b>Method 8</b>										
InHRindir6	25%	44982	-0.48340	0.04839	0.20960	0.00104	0.20757	0.21163	0.09232	0.30385
InHRindir6	50%	45000	-0.46731	0.04161	0.22569	0.00096	0.22380	0.22757	0.09254	0.30421
InHRindir6	75%	45000	-0.45606	0.03892	0.23694	0.00093	0.23512	0.23876	0.09506	0.30831
<b>Method 9</b>										
InHRindir7	25%	44982	-0.54151	0.06126	0.15149	0.00117	0.14921	0.15378	0.08421	0.29018
InHRindir7	50%	45000	-0.52362	0.05250	0.16938	0.00108	0.16726	0.17149	0.08119	0.28494
InHRindir7	75%	45000	-0.51120	0.04935	0.18180	0.00105	0.17974	0.18385	0.08240	0.28705

– Anhang –

Bias for Methods 1 through 9 with HR true = 0.5 [ $\log(\text{HR}_{\text{true}}) = -0.693$ ]

Method	Censoring Distribution	Iterations	Theta	Variance	Bias	MCSE	Lower	Upper	MSE	RMSE
<b>COX PH</b>										
InHRobs	none	45000	-0.69885	0.02604	-0.00585	0.00076	-0.00734	-0.00436	0.02607	0.16146
InHRobs	exp0.5	44982	-0.70212	0.07407	-0.00912	0.00128	-0.01164	-0.00661	0.07415	0.27230
InHRobs	unif0.5	45000	-0.78033	0.05521	-0.08733	0.00111	-0.08950	-0.08516	0.06284	0.25068
<b>Method 1</b>										
InHRdir1	none	45000	-0.63596	0.01691	0.05704	0.00061	0.05584	0.05825	0.02017	0.14201
InHRdir1	exp0.5	44982	-0.67950	0.06476	0.01350	0.00120	0.01114	0.01585	0.06494	0.25484
InHRdir1	unif0.5	45000	-0.70362	0.03587	-0.01062	0.00089	-0.01237	-0.00887	0.03598	0.18970
<b>Method 2</b>										
InHRdir2	none	45000	-0.74386	0.05026	-0.05086	0.00106	-0.05293	-0.04879	0.05284	0.22988
InHRdir2	exp0.5	44982	-0.71194	0.09242	-0.01894	0.00143	-0.02175	-0.01614	0.09277	0.30459
InHRdir2	unif0.5	45000	-0.84064	0.09996	-0.14764	0.00149	-0.15056	-0.14472	0.12176	0.34894
<b>Method 7</b>										
InHRindir5	none	45000	-0.44082	0.03828	0.25218	0.00092	0.25037	0.25399	0.10187	0.31917
InHRindir5	exp0.5	44982	-0.59532	0.07656	0.09768	0.00130	0.09513	0.10024	0.08610	0.29343
InHRindir5	unif0.5	45000	-0.56146	0.05421	0.13154	0.00110	0.12939	0.13369	0.07151	0.26741
<b>Method 8</b>										
InHRindir6	none	45000	-0.39324	0.03039	0.29976	0.00082	0.29815	0.30137	0.12024	0.34676
InHRindir6	exp0.5	44982	-0.51452	0.04815	0.17848	0.00103	0.17645	0.18050	0.08001	0.28286
InHRindir6	unif0.5	45000	-0.49902	0.04204	0.19398	0.00097	0.19209	0.19588	0.07967	0.28226
<b>Method 9</b>										
InHRindir7	none	45000	-0.44082	0.03827	0.25218	0.00092	0.25037	0.25399	0.10187	0.31917
InHRindir7	exp0.5	44982	-0.57631	0.06131	0.11669	0.00117	0.11440	0.11898	0.07493	0.27373
InHRindir7	unif0.5	45000	-0.55922	0.05310	0.13378	0.00109	0.13166	0.13591	0.07099	0.26645

– Anhang –

Bias for Methods 1 through 9 with HR true = 0.5 [ $\log(\text{HR}_{\text{true}}) = -0.693$ ]

Method	Trial Size	Iterations	Theta	Variance	Bias	MCSE	Lower	Upper	MSE	RMSE
<b>COX PH</b>										
InHRobs	N=100	44982	-0.73731	0.12453	-0.04431	0.00166	-0.04757	-0.04105	0.12649	0.35566
InHRobs	N=500	45000	-0.72336	0.02279	-0.03036	0.00071	-0.03175	-0.02896	0.02371	0.15397
InHRobs	N=1000	45000	-0.72066	0.01211	-0.02766	0.00052	-0.02867	-0.02664	0.01288	0.11348
<b>Method 1</b>										
InHRdir1	N=100	44982	-0.66648	0.09255	0.02652	0.00143	0.02371	0.02933	0.09325	0.30537
InHRdir1	N=500	45000	-0.67601	0.01783	0.01699	0.00063	0.01576	0.01822	0.01812	0.13462
InHRdir1	N=1000	45000	-0.67659	0.00946	0.01641	0.00046	0.01552	0.01731	0.00973	0.09865
<b>Method 2</b>										
InHRdir2	N=100	44982	-0.77658	0.17819	-0.08358	0.00199	-0.08748	-0.07968	0.18518	0.43032
InHRdir2	N=500	45000	-0.76131	0.04376	-0.06831	0.00099	-0.07025	-0.06638	0.04842	0.22005
InHRdir2	N=1000	45000	-0.75857	0.02951	-0.06557	0.00081	-0.06716	-0.06398	0.03381	0.18387
<b>Method 7</b>										
InHRindir5	N=100	44982	-0.71006	0.10226	-0.01706	0.00151	-0.02002	-0.01411	0.10255	0.32023
InHRindir5	N=500	45000	-0.51247	0.01470	0.18053	0.00057	0.17941	0.18165	0.04729	0.21746
InHRindir5	N=1000	45000	-0.37512	0.00860	0.31788	0.00044	0.31703	0.31874	0.10964	0.33113
<b>Method 8</b>										
InHRindir6	N=100	44982	-0.62353	0.07149	0.06947	0.00126	0.06700	0.07194	0.07632	0.27626
InHRindir6	N=500	45000	-0.45246	0.00935	0.24054	0.00046	0.23965	0.24144	0.06721	0.25924
InHRindir6	N=1000	45000	-0.33084	0.00524	0.36216	0.00034	0.36149	0.36283	0.13640	0.36932
<b>Method 9</b>										
InHRindir7	N=100	44982	-0.69674	0.08925	-0.00374	0.00141	-0.00650	-0.00098	0.08926	0.29876
InHRindir7	N=500	45000	-0.50772	0.01327	0.18528	0.00054	0.18422	0.18635	0.04760	0.21818
InHRindir7	N=1000	45000	-0.37193	0.00785	0.32107	0.00042	0.32025	0.32189	0.11094	0.33307

– Anhang –

Bias for Methods 1 through 9 with HR true = 0.7 [log(HR\_true) = -0.357]

Method	Allocation Ratio	Iterations	Theta	Variance	Bias	MCSE	Lower	Upper	MSE	RMSE
<b>COX PH</b>										
InHRobs	4:1	26999	-0.37526	0.06065	-0.01826	0.00150	-0.02119	-0.01532	0.06098	0.24694
InHRobs	2:1	27000	-0.37431	0.04162	-0.01731	0.00124	-0.01975	-0.01488	0.04192	0.20475
InHRobs	1:1	27000	-0.37406	0.03484	-0.01706	0.00114	-0.01929	-0.01484	0.03513	0.18744
InHRobs	1:2	27000	-0.37459	0.03939	-0.01759	0.00121	-0.01995	-0.01522	0.03970	0.19925
InHRobs	1:4	26999	-0.37691	0.05511	-0.01991	0.00143	-0.02271	-0.01711	0.05550	0.23559
<b>Method 1</b>										
InHRdir1	4:1	26999	-0.35677	0.05085	0.00023	0.00137	-0.00246	0.00292	0.05085	0.22549
InHRdir1	2:1	27000	-0.35671	0.03473	0.00029	0.00113	-0.00193	0.00251	0.03473	0.18637
InHRdir1	1:1	27000	-0.35797	0.02948	-0.00097	0.00104	-0.00302	0.00108	0.02948	0.17171
InHRdir1	1:2	27000	-0.36116	0.03421	-0.00416	0.00113	-0.00636	-0.00195	0.03423	0.18501
InHRdir1	1:4	26999	-0.36607	0.04904	-0.00907	0.00135	-0.01171	-0.00643	0.04912	0.22163
<b>Method 2</b>										
InHRdir2	4:1	26999	-0.32848	0.04115	0.02852	0.00123	0.02610	0.03094	0.04196	0.20485
InHRdir2	2:1	27000	-0.35245	0.03386	0.00455	0.00112	0.00236	0.00675	0.03388	0.18406
InHRdir2	1:1	27000	-0.38015	0.03656	-0.02315	0.00116	-0.02543	-0.02087	0.03709	0.19259
InHRdir2	1:2	27000	-0.41092	0.05308	-0.05392	0.00140	-0.05666	-0.05117	0.05599	0.23661
InHRdir2	1:4	26999	-0.44567	0.09291	-0.08867	0.00186	-0.09230	-0.08503	0.10076	0.31743
<b>Method 7</b>										
InHRindir5	4:1	26999	-0.37631	0.04414	-0.01931	0.00128	-0.02182	-0.01681	0.04452	0.21099
InHRindir5	2:1	27000	-0.35527	0.03105	0.00173	0.00107	-0.00037	0.00383	0.03106	0.17622
InHRindir5	1:1	27000	-0.34805	0.02562	0.00895	0.00097	0.00704	0.01086	0.02570	0.16030
InHRindir5	1:2	27000	-0.35596	0.02831	0.00104	0.00102	-0.00097	0.00304	0.02831	0.16826
InHRindir5	1:4	26999	-0.37777	0.03823	-0.02077	0.00119	-0.02311	-0.01844	0.03866	0.19661
<b>Method 8</b>										
InHRindir6	4:1	26999	-0.28850	0.02017	0.06850	0.00086	0.06681	0.07019	0.02486	0.15768
InHRindir6	2:1	27000	-0.32678	0.02270	0.03022	0.00092	0.02842	0.03202	0.02361	0.15366
InHRindir6	1:1	27000	-0.34620	0.02447	0.01080	0.00095	0.00893	0.01266	0.02458	0.15679
InHRindir6	1:2	27000	-0.33917	0.02677	0.01783	0.00100	0.01587	0.01978	0.02709	0.16458
InHRindir6	1:4	26999	-0.31004	0.02836	0.04696	0.00102	0.04495	0.04897	0.03056	0.17483
<b>Method 9</b>										
InHRindir7	4:1	26999	-0.36062	0.03152	-0.00362	0.00108	-0.00574	-0.00151	0.03153	0.17757
InHRindir7	2:1	27000	-0.34689	0.02565	0.01011	0.00097	0.00820	0.01202	0.02575	0.16047
InHRindir7	1:1	27000	-0.34620	0.02447	0.01080	0.00095	0.00893	0.01266	0.02458	0.15679
InHRindir7	1:2	27000	-0.36006	0.03026	-0.00306	0.00106	-0.00513	-0.00098	0.03026	0.17397
InHRindir7	1:4	26999	-0.38755	0.04431	-0.03055	0.00128	-0.03306	-0.02804	0.04524	0.21271

– Anhang –

Bias for Methods 1 through 9 with HR true = 0.7 [ $\log(\text{HR}_{\text{true}}) = -0.357$ ]

Method	Baseline		Theta	Variance	Bias	MCSE	Lower	Upper	MSE	RMSE
	Hazard	Iterations								
<b>COX PH</b>										
InHRobs	25%	44998	-0.37598	0.05403	-0.01898	0.00110	-0.02113	-0.01683	0.05439	0.23322
InHRobs	50%	45000	-0.37413	0.04410	-0.01713	0.00099	-0.01907	-0.01519	0.04440	0.21071
InHRobs	75%	45000	-0.37497	0.04083	-0.01797	0.00095	-0.01984	-0.01610	0.04115	0.20286
<b>Method 1</b>										
InHRdir1	25%	44998	-0.36109	0.04719	-0.00409	0.00102	-0.00610	-0.00209	0.04721	0.21727
InHRdir1	50%	45000	-0.35885	0.03751	-0.00185	0.00091	-0.00364	-0.00006	0.03751	0.19369
InHRdir1	75%	45000	-0.35926	0.03432	-0.00226	0.00087	-0.00397	-0.00055	0.03433	0.18528
<b>Method 2</b>										
InHRdir2	25%	44998	-0.38350	0.05967	-0.02650	0.00115	-0.02875	-0.02424	0.06037	0.24571
InHRdir2	50%	45000	-0.38269	0.05142	-0.02569	0.00107	-0.02778	-0.02359	0.05207	0.22820
InHRdir2	75%	45000	-0.38442	0.04861	-0.02742	0.00104	-0.02945	-0.02538	0.04936	0.22218
<b>Method 7</b>										
InHRindir5	25%	44998	-0.37009	0.03971	-0.01309	0.00094	-0.01493	-0.01125	0.03989	0.19971
InHRindir5	50%	45000	-0.36021	0.03204	-0.00321	0.00084	-0.00487	-0.00156	0.03205	0.17902
InHRindir5	75%	45000	-0.35772	0.02901	-0.00072	0.00080	-0.00230	0.00085	0.02901	0.17032
<b>Method 8</b>										
InHRindir6	25%	44998	-0.32750	0.02818	0.02950	0.00079	0.02795	0.03105	0.02905	0.17045
InHRindir6	50%	45000	-0.32015	0.02405	0.03685	0.00073	0.03541	0.03828	0.02540	0.15939
InHRindir6	75%	45000	-0.31877	0.02251	0.03823	0.00071	0.03685	0.03962	0.02397	0.15482
<b>Method 9</b>										
InHRindir7	25%	44998	-0.36628	0.03556	-0.00928	0.00089	-0.01102	-0.00754	0.03565	0.18880
InHRindir7	50%	45000	-0.35813	0.03046	-0.00113	0.00082	-0.00274	0.00048	0.03046	0.17454
InHRindir7	75%	45000	-0.35639	0.02831	0.00061	0.00079	-0.00094	0.00217	0.02831	0.16827

– Anhang –

Bias for Methods 1 through 9 with HR true = 0.7 [ $\log(\text{HR}_{\text{true}}) = -0.357$ ]

Method	Censoring Distribution	Iterations	Theta	Variance	Bias	MCSE	Lower	Upper	MSE	RMSE
<b>COX PH</b>										
InHRobs	none	45000	-0.35979	0.02450	-0.00279	0.00074	-0.00423	-0.00134	0.02451	0.15656
InHRobs	exp0.5	44998	-0.36448	0.06172	-0.00748	0.00117	-0.00977	-0.00518	0.06177	0.24854
InHRobs	unif0.5	45000	-0.40081	0.05174	-0.04381	0.00107	-0.04591	-0.04171	0.05366	0.23165
<b>Method 1</b>										
InHRdir1	none	45000	-0.34408	0.02026	0.01292	0.00067	0.01161	0.01424	0.02042	0.14291
InHRdir1	exp0.5	44998	-0.35562	0.05619	0.00138	0.00112	-0.00081	0.00357	0.05619	0.23704
InHRdir1	unif0.5	45000	-0.37950	0.04193	-0.02250	0.00097	-0.02440	-0.02061	0.04244	0.20600
<b>Method 2</b>										
InHRdir2	none	45000	-0.36960	0.03028	-0.01260	0.00082	-0.01421	-0.01099	0.03044	0.17446
InHRdir2	exp0.5	44998	-0.36716	0.06541	-0.01016	0.00121	-0.01253	-0.00780	0.06551	0.25595
InHRdir2	unif0.5	45000	-0.41383	0.06264	-0.05683	0.00118	-0.05914	-0.05452	0.06587	0.25664
<b>Method 7</b>										
InHRindir5	none	45000	-0.32216	0.01791	0.03484	0.00063	0.03360	0.03608	0.01912	0.13829
InHRindir5	exp0.5	44998	-0.37649	0.04599	-0.01949	0.00101	-0.02147	-0.01751	0.04637	0.21535
InHRindir5	unif0.5	45000	-0.38938	0.03439	-0.03238	0.00087	-0.03409	-0.03066	0.03544	0.18826
<b>Method 8</b>										
InHRindir6	none	45000	-0.28788	0.01422	0.06912	0.00056	0.06801	0.07022	0.01900	0.13783
InHRindir6	exp0.5	44998	-0.33156	0.03200	0.02544	0.00084	0.02379	0.02709	0.03264	0.18067
InHRindir6	unif0.5	45000	-0.34697	0.02669	0.01003	0.00077	0.00852	0.01153	0.02679	0.16367
<b>Method 9</b>										
InHRindir7	none	45000	-0.32216	0.01791	0.03484	0.00063	0.03360	0.03608	0.01912	0.13829
InHRindir7	exp0.5	44998	-0.37075	0.04056	-0.01375	0.00095	-0.01561	-0.01189	0.04075	0.20185
InHRindir7	unif0.5	45000	-0.38789	0.03360	-0.03089	0.00086	-0.03258	-0.02920	0.03456	0.18589

– Anhang –

Bias for Methods 1 through 9 with HR true = 0.7 [ $\log(\text{HR}_{\text{true}}) = -0.357$ ]

Method	Trial Size	Iterations	Theta	Variance	Bias	MCSE	Lower	Upper	MSE	RMSE
<b>COX PH</b>										
InHRobs	N=100	44998	-0.38049	0.10886	-0.02349	0.00156	-0.02654	-0.02044	0.10941	0.33078
InHRobs	N=500	45000	-0.37337	0.02003	-0.01637	0.00067	-0.01768	-0.01507	0.02030	0.14249
InHRobs	N=1000	45000	-0.37122	0.01003	-0.01422	0.00047	-0.01514	-0.01329	0.01023	0.10114
<b>Method 1</b>										
InHRdir1	N=100	44998	-0.35512	0.09181	0.00188	0.00143	-0.00092	0.00468	0.09181	0.30301
InHRdir1	N=500	45000	-0.36204	0.01808	-0.00504	0.00063	-0.00628	-0.00380	0.01810	0.13455
InHRdir1	N=1000	45000	-0.36204	0.00911	-0.00504	0.00045	-0.00592	-0.00416	0.00913	0.09557
<b>Method 2</b>										
InHRdir2	N=100	44998	-0.39171	0.12394	-0.03471	0.00166	-0.03796	-0.03146	0.12515	0.35376
InHRdir2	N=500	45000	-0.38096	0.02351	-0.02396	0.00072	-0.02538	-0.02255	0.02409	0.15520
InHRdir2	N=1000	45000	-0.37792	0.01214	-0.02092	0.00052	-0.02194	-0.01990	0.01258	0.11215
<b>Method 7</b>										
InHRindir5	N=100	44998	-0.40834	0.07523	-0.05134	0.00129	-0.05388	-0.04881	0.07786	0.27904
InHRindir5	N=500	45000	-0.35887	0.01583	-0.00187	0.00059	-0.00303	-0.00071	0.01583	0.12583
InHRindir5	N=1000	45000	-0.32082	0.00594	0.03618	0.00036	0.03547	0.03689	0.00725	0.08512
<b>Method 8</b>										
InHRindir6	N=100	44998	-0.35996	0.05492	-0.00296	0.00110	-0.00513	-0.00080	0.05493	0.23436
InHRindir6	N=500	45000	-0.32057	0.01263	0.03643	0.00053	0.03539	0.03747	0.01396	0.11816
InHRindir6	N=1000	45000	-0.28589	0.00448	0.07111	0.00032	0.07049	0.07173	0.00954	0.09767
<b>Method 9</b>										
InHRindir7	N=100	44998	-0.40284	0.06947	-0.04584	0.00124	-0.04828	-0.04341	0.07157	0.26753
InHRindir7	N=500	45000	-0.35788	0.01565	-0.00088	0.00059	-0.00204	0.00027	0.01565	0.12511
InHRindir7	N=1000	45000	-0.32007	0.00584	0.03693	0.00036	0.03622	0.03764	0.00720	0.08486



– Anhang –

Bias for Methods 1 through 9 with HR true = 0.9 [log(HR\_true) = -0.105]

Method	Allocation Ratio	Iterations	Theta	Variance	Bias	MCSE	Lower	Upper	MSE	RMSE
<b>COX PH</b>										
InHRobs	4:1	27000	-0.10734	0.05541	-0.00234	0.00143	-0.00515	0.00047	0.05541	0.23539
InHRobs	2:1	27000	-0.10928	0.03847	-0.00428	0.00119	-0.00662	-0.00194	0.03849	0.19618
InHRobs	1:1	27000	-0.10957	0.03365	-0.00457	0.00112	-0.00676	-0.00238	0.03367	0.18350
InHRobs	1:2	27000	-0.11238	0.03726	-0.00738	0.00117	-0.00968	-0.00507	0.03732	0.19318
InHRobs	1:4	26999	-0.11290	0.05346	-0.00790	0.00141	-0.01066	-0.00514	0.05352	0.23134
<b>Method 1</b>										
InHRdir1	4:1	27000	-0.10348	0.05013	0.00152	0.00136	-0.00115	0.00420	0.05013	0.22389
InHRdir1	2:1	27000	-0.10576	0.03471	-0.00076	0.00113	-0.00299	0.00146	0.03471	0.18630
InHRdir1	1:1	27000	-0.10657	0.03041	-0.00157	0.00106	-0.00365	0.00051	0.03041	0.17439
InHRdir1	1:2	27000	-0.10969	0.03393	-0.00469	0.00112	-0.00688	-0.00249	0.03395	0.18425
InHRdir1	1:4	26999	-0.11080	0.04884	-0.00580	0.00134	-0.00844	-0.00317	0.04887	0.22106
<b>Method 2</b>										
InHRdir2	4:1	27000	-0.08952	0.05200	0.01548	0.00139	0.01276	0.01820	0.05224	0.22855
InHRdir2	2:1	27000	-0.10199	0.03641	0.00301	0.00116	0.00073	0.00528	0.03642	0.19083
InHRdir2	1:1	27000	-0.11026	0.03417	-0.00526	0.00112	-0.00746	-0.00306	0.03419	0.18492
InHRdir2	1:2	27000	-0.12141	0.04116	-0.01641	0.00123	-0.01883	-0.01399	0.04143	0.20355
InHRdir2	1:4	26999	-0.13398	0.06542	-0.02898	0.00156	-0.03203	-0.02593	0.06626	0.25741
<b>Method 7</b>										
InHRindir5	4:1	27000	-0.18689	0.03071	-0.08189	0.00107	-0.08398	-0.07980	0.03742	0.19343
InHRindir5	2:1	27000	-0.16608	0.02126	-0.06108	0.00089	-0.06282	-0.05934	0.02499	0.15807
InHRindir5	1:1	27000	-0.16031	0.01845	-0.05531	0.00083	-0.05693	-0.05369	0.02151	0.14665
InHRindir5	1:2	27000	-0.16680	0.02058	-0.06180	0.00087	-0.06351	-0.06009	0.02439	0.15619
InHRindir5	1:4	26999	-0.18727	0.02972	-0.08227	0.00105	-0.08433	-0.08022	0.03649	0.19101
<b>Method 8</b>										
InHRindir6	4:1	27000	-0.14576	0.01660	-0.04076	0.00078	-0.04229	-0.03922	0.01826	0.13513
InHRindir6	2:1	27000	-0.15432	0.01722	-0.04932	0.00080	-0.05089	-0.04776	0.01965	0.14017
InHRindir6	1:1	27000	-0.15968	0.01799	-0.05468	0.00082	-0.05628	-0.05308	0.02098	0.14484
InHRindir6	1:2	27000	-0.15757	0.01850	-0.05257	0.00083	-0.05419	-0.05095	0.02126	0.14582
InHRindir6	1:4	26999	-0.15072	0.01953	-0.04572	0.00085	-0.04739	-0.04406	0.02162	0.14703
<b>Method 9</b>										
InHRindir7	4:1	27000	-0.18220	0.02594	-0.07720	0.00098	-0.07912	-0.07527	0.03189	0.17859
InHRindir7	2:1	27000	-0.16388	0.01946	-0.05888	0.00085	-0.06054	-0.05721	0.02292	0.15140
InHRindir7	1:1	27000	-0.15968	0.01799	-0.05468	0.00082	-0.05628	-0.05308	0.02098	0.14484
InHRindir7	1:2	27000	-0.16732	0.02091	-0.06232	0.00088	-0.06405	-0.06060	0.02479	0.15745
InHRindir7	1:4	26999	-0.18840	0.03051	-0.08340	0.00106	-0.08549	-0.08132	0.03747	0.19357

– Anhang –

Bias for Methods 1 through 9 with HR true = 0.9 [log(HR\_true) = -0.105]

Method	Baseline Hazard	Iterations	Theta	Variance	Bias	MCSE	Lower	Upper	MSE	RMSE
<b>COX PH</b>										
InHRobs	25%	44999	-0.11097	0.05108	-0.00597	0.00107	-0.00806	-0.00388	0.05111	0.22608
InHRobs	50%	45000	-0.11076	0.04141	-0.00576	0.00096	-0.00764	-0.00388	0.04144	0.20358
InHRobs	75%	45000	-0.10915	0.03847	-0.00415	0.00092	-0.00596	-0.00233	0.03849	0.19618
<b>Method 1</b>										
InHRdir1	25%	44999	-0.10797	0.04665	-0.00297	0.00102	-0.00497	-0.00098	0.04666	0.21601
InHRdir1	50%	45000	-0.10762	0.03743	-0.00262	0.00091	-0.00440	-0.00083	0.03744	0.19349
InHRdir1	75%	45000	-0.10618	0.03474	-0.00118	0.00088	-0.00291	0.00054	0.03474	0.18639
<b>Method 2</b>										
InHRdir2	25%	44999	-0.11181	0.05278	-0.00681	0.00108	-0.00893	-0.00469	0.05282	0.22983
InHRdir2	50%	45000	-0.11196	0.04435	-0.00696	0.00099	-0.00890	-0.00501	0.04439	0.21070
InHRdir2	75%	45000	-0.11053	0.04108	-0.00553	0.00096	-0.00740	-0.00365	0.04111	0.20275
<b>Method 7</b>										
InHRindir5	25%	44999	-0.18252	0.02908	-0.07752	0.00080	-0.07910	-0.07594	0.03508	0.18731
InHRindir5	50%	45000	-0.17097	0.02278	-0.06597	0.00071	-0.06736	-0.06457	0.02713	0.16471
InHRindir5	75%	45000	-0.16693	0.02082	-0.06193	0.00068	-0.06326	-0.06059	0.02466	0.15703
<b>Method 8</b>										
InHRindir6	25%	44999	-0.16090	0.02070	-0.05590	0.00068	-0.05723	-0.05457	0.02382	0.15434
InHRindir6	50%	45000	-0.15162	0.01722	-0.04662	0.00062	-0.04784	-0.04541	0.01940	0.13927
InHRindir6	75%	45000	-0.14831	0.01597	-0.04331	0.00060	-0.04448	-0.04214	0.01785	0.13359
<b>Method 9</b>										
InHRindir7	25%	44999	-0.18050	0.02655	-0.07550	0.00077	-0.07701	-0.07400	0.03225	0.17957
InHRindir7	50%	45000	-0.17009	0.02217	-0.06509	0.00070	-0.06647	-0.06372	0.02641	0.16250
InHRindir7	75%	45000	-0.16629	0.02042	-0.06129	0.00067	-0.06261	-0.05997	0.02418	0.15550

– Anhang –

Bias for Methods 1 through 9 with HR true = 0.9 [log(HR\_true) = -0.105]

Method	Censoring Distribution	Iterations	Theta	Variance	Bias	MCSE	Lower	Upper	MSE	RMSE
<b>COX PH</b>										
InHRobs	none	45000	-0.10641	0.02388	-0.00141	0.00073	-0.00284	0.00001	0.02388	0.15455
InHRobs	exp0.5	44999	-0.10510	0.05690	-0.00010	0.00112	-0.00230	0.00211	0.05690	0.23853
InHRobs	unif0.5	45000	-0.11937	0.05006	-0.01437	0.00105	-0.01643	-0.01230	0.05026	0.22419
<b>Method 1</b>										
InHRdir1	none	45000	-0.10372	0.02176	0.00128	0.00070	-0.00008	0.00264	0.02176	0.14751
InHRdir1	exp0.5	44999	-0.10296	0.05283	0.00204	0.00108	-0.00008	0.00416	0.05284	0.22986
InHRdir1	unif0.5	45000	-0.11509	0.04414	-0.01009	0.00099	-0.01204	-0.00815	0.04424	0.21034
<b>Method 2</b>										
InHRdir2	none	45000	-0.10781	0.02601	-0.00281	0.00076	-0.00430	-0.00132	0.02601	0.16129
InHRdir2	exp0.5	44999	-0.10548	0.05801	-0.00048	0.00114	-0.00270	0.00175	0.05801	0.24085
InHRdir2	unif0.5	45000	-0.12101	0.05404	-0.01601	0.00110	-0.01815	-0.01386	0.05430	0.23302
<b>Method 7</b>										
InHRindir5	none	45000	-0.14438	0.01318	-0.03938	0.00054	-0.04044	-0.03832	0.01473	0.12135
InHRindir5	exp0.5	44999	-0.18902	0.03185	-0.08402	0.00084	-0.08567	-0.08237	0.03891	0.19725
InHRindir5	unif0.5	45000	-0.18702	0.02651	-0.08202	0.00077	-0.08352	-0.08051	0.03324	0.18231
<b>Method 8</b>										
InHRindir6	none	45000	-0.12885	0.01033	-0.02385	0.00048	-0.02479	-0.02291	0.01090	0.10440
InHRindir6	exp0.5	44999	-0.16630	0.02260	-0.06130	0.00071	-0.06269	-0.05991	0.02635	0.16233
InHRindir6	unif0.5	45000	-0.16568	0.02013	-0.06068	0.00067	-0.06199	-0.05937	0.02381	0.15431
<b>Method 9</b>										
InHRindir7	none	45000	-0.14438	0.01318	-0.03938	0.00054	-0.04044	-0.03832	0.01473	0.12135
InHRindir7	exp0.5	44999	-0.18664	0.02902	-0.08164	0.00080	-0.08321	-0.08006	0.03569	0.18890
InHRindir7	unif0.5	45000	-0.18587	0.02588	-0.08087	0.00076	-0.08236	-0.07938	0.03242	0.18006

– Anhang –

Bias for Methods 1 through 9 with HR true = 0.9 [ $\log(\text{HR}_{\text{true}}) = -0.105$ ]

Method	Trial Size	Iterations	Theta	Variance	Bias	MCSE	Lower	Upper	MSE	RMSE
<b>COX PH</b>										
InHRobs	N=100	44999	-0.11197	0.10320	-0.00697	0.00151	-0.00994	-0.00400	0.10325	0.32133
InHRobs	N=500	45000	-0.10935	0.01862	-0.00435	0.00064	-0.00561	-0.00309	0.01864	0.13654
InHRobs	N=1000	45000	-0.10956	0.00913	-0.00456	0.00045	-0.00544	-0.00368	0.00915	0.09566
<b>Method 1</b>										
InHRdir1	N=100	44999	-0.10575	0.09195	-0.00075	0.00143	-0.00355	0.00206	0.09195	0.30324
InHRdir1	N=500	45000	-0.10756	0.01796	-0.00256	0.00063	-0.00380	-0.00132	0.01796	0.13403
InHRdir1	N=1000	45000	-0.10847	0.00891	-0.00347	0.00045	-0.00434	-0.00260	0.00892	0.09446
<b>Method 2</b>										
InHRdir2	N=100	44999	-0.11427	0.10987	-0.00927	0.00156	-0.01233	-0.00620	0.10995	0.33159
InHRdir2	N=500	45000	-0.10999	0.01902	-0.00499	0.00065	-0.00627	-0.00372	0.01904	0.13799
InHRdir2	N=1000	45000	-0.11004	0.00930	-0.00504	0.00045	-0.00593	-0.00415	0.00933	0.09659
<b>Method 7</b>										
InHRindir5	N=100	44999	-0.25904	0.04477	-0.15404	0.00100	-0.15599	-0.15208	0.06849	0.26171
InHRindir5	N=500	45000	-0.14029	0.01056	-0.03529	0.00048	-0.03624	-0.03434	0.01180	0.10864
InHRindir5	N=1000	45000	-0.12109	0.00632	-0.01609	0.00037	-0.01682	-0.01535	0.00658	0.08110
<b>Method 8</b>										
InHRindir6	N=100	44999	-0.22766	0.03236	-0.12266	0.00085	-0.12432	-0.12099	0.04740	0.21772
InHRindir6	N=500	45000	-0.12499	0.00824	-0.01999	0.00043	-0.02083	-0.01915	0.00864	0.09295
InHRindir6	N=1000	45000	-0.10819	0.00501	-0.00319	0.00033	-0.00384	-0.00254	0.00502	0.07086
<b>Method 9</b>										
InHRindir7	N=100	44999	-0.25595	0.04184	-0.15095	0.00096	-0.15284	-0.14907	0.06463	0.25422
InHRindir7	N=500	45000	-0.13997	0.01044	-0.03497	0.00048	-0.03591	-0.03402	0.01167	0.10801
InHRindir7	N=1000	45000	-0.12097	0.00629	-0.01597	0.00037	-0.01670	-0.01524	0.00654	0.08089

– Anhang –

Pooled Variance Estimators for Methods 1 through 6 at differing Allocation Ratios

Method	Allocation Ratio	Iterations	Theta	Variance	MCSE	Lower	Upper
<b>COX PH</b>							
InVobs	4:1	134189	0.08804	0.02110	0.00040	0.08726	0.08881
InVobs	2:1	134673	0.06202	0.01150	0.00029	0.06145	0.06260
InVobs	1:1	134875	0.05107	0.00743	0.00023	0.05061	0.05153
InVobs	1:2	134952	0.05179	0.00618	0.00021	0.05137	0.05220
InVobs	1:4	134953	0.06503	0.00816	0.00025	0.06455	0.06551
<b>Method 1</b>							
InVdir1	4:1	134189	0.04045	0.00228	0.00013	0.04019	0.04070
InVdir1	2:1	134673	0.03455	0.00152	0.00011	0.03435	0.03476
InVdir1	1:1	134875	0.03876	0.00196	0.00012	0.03852	0.03899
InVdir1	1:2	134952	0.05761	0.00521	0.00020	0.05723	0.05800
InVdir1	1:4	134953	0.10385	0.02186	0.00040	0.10306	0.10464
<b>Method 2</b>							
InVdir2	4:1	134189	0.04423	0.00257	0.00014	0.04396	0.04451
InVdir2	2:1	134673	0.03796	0.00179	0.00012	0.03774	0.03819
InVdir2	1:1	134875	0.04230	0.00237	0.00013	0.04204	0.04256
InVdir2	1:2	134952	0.06181	0.00617	0.00021	0.06139	0.06223
InVdir2	1:4	134953	0.10927	0.02474	0.00043	0.10843	0.11011
<b>Method 3</b>							
InVindir1	4:1	134189	0.08600	0.01820	0.00037	0.08527	0.08672
InVindir1	2:1	134673	0.06086	0.00976	0.00027	0.06033	0.06138
InVindir1	1:1	134875	0.05015	0.00598	0.00021	0.04973	0.05056
InVindir1	1:2	134952	0.05125	0.00529	0.00020	0.05086	0.05164
InVindir1	1:4	134953	0.06449	0.00729	0.00023	0.06404	0.06495
<b>Method 4</b>							
InVindir2	4:1	134189	0.07019	0.01478	0.00033	0.06954	0.07084
InVindir2	2:1	134673	0.04947	0.00815	0.00025	0.04898	0.04995
InVindir2	1:1	134875	0.04115	0.00519	0.00020	0.04077	0.04154
InVindir2	1:2	134952	0.04274	0.00424	0.00018	0.04239	0.04309
InVindir2	1:4	134953	0.05566	0.00577	0.00021	0.05525	0.05606

**Method 5**

InVindir3	4:1	134189	0.03103	0.00113	0.00009	0.03085	0.03121
InVindir3	2:1	134673	0.03233	0.00129	0.00010	0.03214	0.03252
InVindir3	1:1	134875	0.03417	0.00158	0.00011	0.03396	0.03438
InVindir3	1:2	134952	0.03669	0.00218	0.00013	0.03644	0.03694
InVindir3	1:4	134953	0.03968	0.00328	0.00016	0.03938	0.03999

**Method 6**

InVindir4	4:1	134189	0.04849	0.00277	0.00014	0.04820	0.04877
InVindir4	2:1	134673	0.03651	0.00165	0.00011	0.03629	0.03672
InVindir4	1:1	134875	0.03417	0.00158	0.00011	0.03396	0.03438
InVindir4	1:2	134952	0.04143	0.00279	0.00014	0.04115	0.04172
InVindir4	1:4	134953	0.06200	0.00801	0.00024	0.06153	0.06248

Pooled Variance Estimators for Methods 1 through 6 at differing Baseline Hazards

Method	Baseline Hazard	Iterations	Theta	Variance	MCSE	Lower	Upper
<b>COX PH</b>							
InVobs	25%	224055	0.07609	0.01664	0.00027	0.07556	0.07663
InVobs	50%	224703	0.06012	0.00912	0.00020	0.05973	0.06052
InVobs	75%	224884	0.05452	0.00713	0.00018	0.05417	0.05487
<b>Method 1</b>							
InVdir1	25%	224055	0.06379	0.01002	0.00021	0.06337	0.06420
InVdir1	50%	224703	0.05259	0.00614	0.00017	0.05226	0.05291
InVdir1	75%	224884	0.04887	0.00542	0.00016	0.04857	0.04918
<b>Method 2</b>							
InVdir2	25%	224055	0.06792	0.01113	0.00022	0.06749	0.06836
InVdir2	50%	224703	0.05662	0.00709	0.00018	0.05627	0.05697
InVdir2	75%	224884	0.05292	0.00637	0.00017	0.05259	0.05325
<b>Method 3</b>							
InVindir1	25%	224055	0.07493	0.01487	0.00026	0.07442	0.07543
InVindir1	50%	224703	0.05914	0.00761	0.00018	0.05877	0.05950
InVindir1	75%	224884	0.05356	0.00569	0.00016	0.05324	0.05387
<b>Method 4</b>							
InVindir2	25%	224055	0.06465	0.01336	0.00024	0.06417	0.06513
InVindir2	50%	224703	0.04837	0.00584	0.00016	0.04806	0.04869
InVindir2	75%	224884	0.04249	0.00375	0.00013	0.04223	0.04274
<b>Method 5</b>							
InVindir3	25%	224055	0.04164	0.00322	0.00012	0.04140	0.04187
InVindir3	50%	224703	0.03287	0.00140	0.00008	0.03271	0.03302
InVindir3	75%	224884	0.02988	0.00103	0.00007	0.02974	0.03001
<b>Method 6</b>							
InVindir4	25%	224055	0.05334	0.00592	0.00016	0.05302	0.05365
InVindir4	50%	224703	0.04205	0.00252	0.00011	0.04184	0.04226
InVindir4	75%	224884	0.03821	0.00184	0.00009	0.03803	0.03838

– Anhang –

Pooled Variance Estimators for Methods 1 through 6 at differing Censoring Distributions

Method	Censoring Distribution	Iterations	Theta	Variance	MCSE	Lower	Upper
<b>COX PH</b>							
InVobs	none	224999	0.03066	0.00135	0.00008	0.03050	0.03081
InVobs	exp0.5	223750	0.09404	0.02178	0.00031	0.09343	0.09465
InVobs	unif0.5	224893	0.06617	0.00804	0.00019	0.06580	0.06654
<b>Method 1</b>							
InVdir1	none	224999	0.02968	0.00158	0.00008	0.02951	0.02984
InVdir1	exp0.5	223750	0.07141	0.00973	0.00021	0.07100	0.07182
InVdir1	unif0.5	224893	0.06423	0.00940	0.00020	0.06383	0.06463
<b>Method 2</b>							
InVdir2	none	224999	0.03271	0.00192	0.00009	0.03253	0.03289
InVdir2	exp0.5	223750	0.07376	0.01037	0.00022	0.07334	0.07418
InVdir2	unif0.5	224893	0.07104	0.01136	0.00022	0.07060	0.07149
<b>Method 3</b>							
InVindir1	none	224999	0.03068	0.00132	0.00008	0.03053	0.03083
InVindir1	exp0.5	223750	0.09217	0.01889	0.00029	0.09160	0.09274
InVindir1	unif0.5	224893	0.06489	0.00634	0.00017	0.06456	0.06521
<b>Method 4</b>							
InVindir2	none	224999	0.02248	0.00047	0.00005	0.02239	0.02257
InVindir2	exp0.5	223750	0.08760	0.01846	0.00029	0.08703	0.08816
InVindir2	unif0.5	224893	0.04558	0.00214	0.00010	0.04539	0.04577
<b>Method 5</b>							
InVindir3	none	224999	0.01758	0.00027	0.00003	0.01751	0.01765
InVindir3	exp0.5	223750	0.05194	0.00378	0.00013	0.05169	0.05220
InVindir3	unif0.5	224893	0.03493	0.00109	0.00007	0.03479	0.03507
<b>Method 6</b>							
InVindir4	none	224999	0.02244	0.00047	0.00005	0.02235	0.02253
InVindir4	exp0.5	223750	0.06665	0.00706	0.00018	0.06630	0.06700
InVindir4	unif0.5	224893	0.04459	0.00190	0.00009	0.04441	0.04477



Pooled Variance Estimators for Methods 1 through 6 at differing Hazard Ratios

Method	Hazard Ratio	Iterations	Theta	Variance	MCSE	Lower	Upper
<b>COX PH</b>							
InVobs	0.1	133747	0.12006	0.03279	0.00050	0.11909	0.12103
InVobs	0.3	134916	0.06259	0.00841	0.00025	0.06210	0.06308
InVobs	0.5	134982	0.04950	0.00440	0.00018	0.04915	0.04986
InVobs	0.7	134998	0.04429	0.00307	0.00015	0.04399	0.04459
InVobs	0.9	134999	0.04189	0.00253	0.00014	0.04162	0.04216
<b>Method 1</b>							
InVdir1	0.1	133747	0.08615	0.01987	0.00039	0.08540	0.08691
InVdir1	0.3	134916	0.05854	0.00661	0.00022	0.05811	0.05898
InVdir1	0.5	134982	0.04773	0.00362	0.00016	0.04741	0.04805
InVdir1	0.7	134998	0.04277	0.00259	0.00014	0.04250	0.04304
InVdir1	0.9	134999	0.04045	0.00218	0.00013	0.04021	0.04070
<b>Method 2</b>							
InVdir2	0.1	133747	0.09588	0.02276	0.00041	0.09507	0.09668
InVdir2	0.3	134916	0.06342	0.00741	0.00023	0.06296	0.06388
InVdir2	0.5	134982	0.05047	0.00399	0.00017	0.05013	0.05080
InVdir2	0.7	134998	0.04450	0.00283	0.00014	0.04422	0.04479
InVdir2	0.9	134999	0.04180	0.00236	0.00013	0.04154	0.04206
<b>Method 3</b>							
InVindir1	0.1	133747	0.11478	0.02539	0.00044	0.11393	0.11563
InVindir1	0.3	134916	0.06262	0.00844	0.00025	0.06213	0.06311
InVindir1	0.5	134982	0.04951	0.00439	0.00018	0.04916	0.04986
InVindir1	0.7	134998	0.04430	0.00307	0.00015	0.04400	0.04459
InVindir1	0.9	134999	0.04190	0.00254	0.00014	0.04163	0.04217
<b>Method 4</b>							
InVindir2	0.1	133747	0.07981	0.02225	0.00041	0.07901	0.08061
InVindir2	0.3	134916	0.05220	0.00671	0.00022	0.05176	0.05264
InVindir2	0.5	134982	0.04507	0.00374	0.00017	0.04474	0.04540
InVindir2	0.7	134998	0.04198	0.00273	0.00014	0.04171	0.04226
InVindir2	0.9	134999	0.04031	0.00231	0.00013	0.04006	0.04057

**Method 5**

InVindir3	0.1	133747	0.04182	0.00371	0.00017	0.04149	0.04214
InVindir3	0.3	134916	0.03611	0.00196	0.00012	0.03588	0.03635
InVindir3	0.5	134982	0.03340	0.00146	0.00010	0.03319	0.03360
InVindir3	0.7	134998	0.03185	0.00123	0.00010	0.03166	0.03204
InVindir3	0.9	134999	0.03082	0.00110	0.00009	0.03064	0.03100

**Method 6**

InVindir4	0.1	133747	0.05382	0.00712	0.00023	0.05337	0.05427
InVindir4	0.3	134916	0.04620	0.00350	0.00016	0.04588	0.04651
InVindir4	0.5	134982	0.04266	0.00254	0.00014	0.04239	0.04292
InVindir4	0.7	134998	0.04067	0.00214	0.00013	0.04042	0.04091
InVindir4	0.9	134999	0.03934	0.00191	0.00012	0.03911	0.03958

Pooled Variance Estimators for Methods 1 through 6 at differing Trial Sizes

Method	Trial Size	Iterations	Theta	Variance	MCSE	Lower	Upper
<b>COX PH</b>							
InVobs	N=100	223649	0.14964	0.02096	0.00031	0.14904	0.15024
InVobs	N=500	224993	0.02793	0.00092	0.00006	0.02780	0.02805
InVobs	N=1000	225000	0.01364	0.00017	0.00003	0.01359	0.01369
<b>Method 1</b>							
InVdir1	N=100	223649	0.12821	0.01315	0.00024	0.12774	0.12869
InVdir1	N=500	224993	0.02498	0.00043	0.00004	0.02489	0.02507
InVdir1	N=1000	225000	0.01246	0.00011	0.00002	0.01242	0.01250
<b>Method 2</b>							
InVdir2	N=100	223649	0.13812	0.01477	0.00026	0.13761	0.13862
InVdir2	N=500	224993	0.02656	0.00048	0.00005	0.02647	0.02665
InVdir2	N=1000	225000	0.01322	0.00012	0.00002	0.01318	0.01327
<b>Method 3</b>							
InVindir1	N=100	223649	0.14647	0.01686	0.00027	0.14593	0.14701
InVindir1	N=500	224993	0.02792	0.00082	0.00006	0.02780	0.02804
InVindir1	N=1000	225000	0.01368	0.00017	0.00003	0.01363	0.01374
<b>Method 4</b>							
InVindir2	N=100	223649	0.12020	0.01520	0.00026	0.11969	0.12071
InVindir2	N=500	224993	0.02392	0.00085	0.00006	0.02380	0.02404
InVindir2	N=1000	225000	0.01176	0.00016	0.00003	0.01171	0.01181
<b>Method 5</b>							
InVindir3	N=100	223649	0.08063	0.00246	0.00010	0.08042	0.08083
InVindir3	N=500	224993	0.01601	0.00008	0.00002	0.01597	0.01605
InVindir3	N=1000	225000	0.00799	0.00002	0.00001	0.00798	0.00801
<b>Method 6</b>							
InVindir4	N=100	223649	0.10323	0.00501	0.00015	0.10293	0.10352
InVindir4	N=500	224993	0.02046	0.00017	0.00003	0.02041	0.02052
InVindir4	N=1000	225000	0.01022	0.00004	0.00001	0.01020	0.01025