

Aus dem Universitätsklinikum Münster
Institut für Biometrie und Klinische Forschung
Geschäftsführender Direktor:
Univ.-Prof. Dr. rer. nat. et med. habil. Andreas Faldum

Bayesianische Verfahren in pädiatrischen Studien

Extrapolation historischer Informationen aus klinischen Studien an Erwachsenen und Kindern: Ergebnisse, Limitationen und Erweiterungen eines meta-analytisch prädiktiven Ansatzes

INAUGURAL - DISSERTATION

zur Erlangung des doctor rerum medicinalium

der Medizinischen Fakultät
der Westfälischen Wilhelms-Universität Münster

vorgelegt von Koch, Raphael
aus Marsberg

2014

Gedruckt mit Genehmigung der Medizinischen Fakultät der Westfälischen
Wilhelms-Universität Münster

Dekan: Univ.-Prof. Dr. med. Dr. h.c. Wilhelm Schmitz

1. Berichterstatter: Univ.-Prof. em. Dr. rer. pol. Wolfgang Köpcke

2. Berichterstatter: Univ.-Prof. Dr. rer. nat. et med. habil. Andreas Faldum

Tag der mündlichen Prüfung: 24.03.2014

Aus dem Universitätsklinikum Münster
Institut für Biometrie und Klinische Forschung
Direktor: Univ.-Prof. Dr. rer. nat. et med. habil. Andreas Faldum
Referent: Univ.-Prof. em. Dr. rer. pol. Wolfgang Köpcke
Koreferent: Univ.-Prof. Dr. rer. nat. et med. habil. Andreas Faldum

Zusammenfassung

Bayesianische Verfahren in pädiatrischen Studien

Extrapolation historischer Informationen aus klinischen Studien an Erwachsenen und Kindern:

Ergebnisse, Limitationen und Erweiterungen eines meta-analytisch prädiktiven Ansatzes

Raphael Koch

In dieser Arbeit wird ein statistisches Bayes-Verfahren mit dem Ziel der Fallzahlreduzierung in einer neu geplanten klinischen Studie untersucht. Um dieses Ziel zu erreichen, wird ein meta-analytisch prädiktiver Bayes-Ansatz angewandt, der Ergebnisse aus historischen Studien an Erwachsenen und Kindern extrapoliert. Diese Methode basiert auf hierarchischen Modellen, die den Informationsgehalt der historischen Studien über einen zukünftigen Therapieerfolg bewerten. Das Ergebnis ist die a-posteriori prädiktive Verteilung über einen zukünftigen studienspezifischen Behandlungseffekt. Damit kann empirisch ein Gewichtungsfaktor bestimmt werden, der die Herabgewichtung der historischen Information in Form der Inversen-Varianzen quantifiziert. Für eine konkrete neue Studie kann die Anzahl der Patienten berechnet werden, die durch Anwendung der a-posteriori prädiktiven Parameter als informative a-priori Parameter bei der Rekrutierung eingespart werden kann. Das Vorwissen wird somit in eine virtuelle Anzahl Patienten, der Prior Effective Sample Size übersetzt. Der entscheidende Parameter zur Bewertung des Extrapolationsumfangs ist die Interstudienvarianz.

Im Rahmen der hierarchischen Modelle wird untersucht, welche hyper a-priori Verteilungen für die Interstudienvarianz am geeignetsten sind. In verschiedenen klinischen Beispielen wird die Effektivität dieser Methodik getestet und bewertet. Insbesondere wird die Situation bei Vorliegen weniger früherer Studien genauer analysiert. In verschiedenen Simulationsszenarien wird die Mindestanzahl historischer Studien bestimmt, ab der die Anwendung dieser Methode bzw. die Extrapolation von historischen Studienresultaten robust und sinnvoll ist.

Ergebnisse: Erst ab fünf Vorstudien entstehen stabile Schätzungen des zukünftigen Studieneffekts. Die geeignetste a-priori Verteilung für die Interstudienvarianz für die hier betrachteten Zielgrößen und Szenarien ist die Half-Normal-Verteilung $HN(0,1)$. Die Anwendung der Inverse-Gamma-Verteilung eignet sich nur bedingt, da sie sich antikonservativ verhält und den Informationsgehalt der historischen Studien überbewertet. Unter Berücksichtigung einer geeigneten Modellanpassung und ausreichender Anzahl historischer Studien kann die Fallzahl einer neuen Studie erheblich reduziert werden. In den klinischen Beispielen beträgt die Einsparung unter Anwendung der $HN(0,1)$ -Verteilung in einer Kinderstudie bis zu 7%, in einer Erwachsenenstudie bis zu 27%. In den Simulationen ist sie teils 100%.

Tag der mündlichen Prüfung: 24.03.2014

Erklärung

Ich gebe hiermit die Erklärung ab, dass ich die Dissertation mit dem Titel
Bayesianische Verfahren in pädiatrischen Studien
Extrapolation historischer Informationen aus klinischen Studien an Erwachsenen und Kindern:
Ergebnisse, Limitationen und Erweiterungen eines meta-analytisch prädiktiven Ansatzes

im

Institut für Biometrie und Klinische Forschung (IBKF)

unter Anleitung von

Univ.-Prof. em. Dr. rer. pol. Wolfgang Köpcke

1. selbständig angefertigt,
2. nur unter Benutzung der im Literaturverzeichnis angegebenen Arbeiten angefertigt und sonst kein anderes gedrucktes oder ungedrucktes Material verwendet,
3. keine unerlaubte fremde Hilfe in Anspruch genommen,
4. sie weder in der gegenwärtigen noch in einer anderen Fassung einer in- oder ausländischen Fakultät als Dissertation, Semesterarbeit, Prüfungsarbeit, oder zur Erlangung eines akademischen Grades, vorgelegt habe.

Ort, Datum

Raphael Koch

Danksagung

An erster Stelle möchte ich mich bei meinen Eltern bedanken, die mich immer unterstützt haben und mich durch ihren Optimismus und ihr Vertrauen stets animiert und wieder aufgebaut haben.

Herrn Professor Wolfgang Köpcke danke ich vielmals für die umfassende Betreuung während meiner Dissertation. Auch über seine Emeritierung hinaus nahm er diese Aufgabe wahr, was ich nicht für selbstverständlich erachte. Mit seinem umfangreichen Fachwissen konnte er mir stets gute Ratschläge geben und die Arbeit damit in die richtigen Bahnen lenken. Dennoch konnte ich ausreichend Kreativität bei der Umsetzung der Dissertation einbringen. Ganz herzlichen Dank für alles.

Herrn Professor Andreas Faldum danke ich für seine detaillierten fachlichen Fragen und vielen konstruktiven Anregungen, die ich gerne in dieser Arbeit umgesetzt habe. Er hatte immer ein offenes Ohr und gab mir durch anderweitige Entlastung die Möglichkeit diese Arbeit zu schreiben. Vielen lieben Dank für die langen und fruchtbaren Diskussionen bis in die späte Nacht hinein.

Herrn Professor Martin Dugas sage ich Dank für die Mitgliedschaft in meinem Dissertationskomitee.

Mein ganz besonderer Dank gilt Joachim Gerß. Ohne ihn wäre diese Arbeit gar nicht möglich gewesen, geschweige denn zustande gekommen. Ich danke ihm für die Vorarbeiten seinerseits und dass ich ihn jederzeit mit methodischen Fragen aufsuchen durfte. Danke, dass du nie die Geduld verloren hast und immer bereit warst methodische Probleme mit mir zu diskutieren.

Weiter danke ich allen, die diese Arbeit tatsächlich durchgelesen und mich auf sprachliche Fehler hingewiesen haben. Tausend Dank.

Zum Schluss möchte ich allen Kollegen danken, die mir bei kleinen Programmierschwierigkeiten unterstützend zur Seite standen und allen Freunden für Ihr Verständnis, obwohl ich sie oft versetzt habe.

Ihnen/Euch allen sage ich ganz herzlich Dank!

Inhaltsverzeichnis

Tabellenverzeichnis	9
Abbildungsverzeichnis	10
Notationen	11
I. Einleitung	
1. Motivation	15
1.1. Fallzahlproblematik	16
1.2. Fallstricke historischer Daten	18
II. Methoden	
2. Grundlagen	23
2.1. Grundlagen der Bayes-Statistik	24
2.2. Normal-Modell	27
2.2.1. Normal-Modell bei einer Beobachtungen	27
2.2.2. Normal-Modell mit mehreren Beobachtungen	29
2.3. Zusammenhänge von Studieneffekten	31
2.3.1. Modellannahmen und Evidenz	31
2.4. Hierarchische Modelle	35
2.4.1. MCMC-Verfahren	37
2.4.2. Normal-Normal-Modell	39
2.5. Auswahl der a-priori Verteilungen	44
2.5.1. Kritik an der Auswahl	44
2.5.2. Hyper a-priori Verteilungen	46
3. Meta-analytisch prädiktiver Ansatz	53
3.1. Historischer Überblick der Methodik	53
3.2. Volles Modell	55
3.3. Gewichtung der Präzision	58
3.3.1. Referenzmodell	59
3.3.2. Precision Weight Effective Sample Size	61
3.3.3. Validierung und Poweranalyse	65
3.4. Prior Effective Sample Size	68
3.5. MCMC-Einstellungen	71

3.6. Getrimmtes Modell	72
3.7. Software	74

III. Ergebnisse

4. Anwendungen	76
4.1. Heparin-Studien	76
4.2. Metformin-Studien	90
4.3. FeNO-Studien	94
5. Simulationsstudie	106
5.1. Settings	107
5.1.1. Szenario A	108
5.1.2. Szenario B	109
5.1.3. Generierung der historischen Studien	109
5.1.4. MCMC-Simulationen	110
5.1.5. Powerbetrachtung	111
5.1.6. Berechnung der Prior Effective Sample Size	111
5.2. Adjustiertes Simulationssetting	112
5.3. Ergebnisse der Simulationsstudie	114
5.3.1. Ergebnisse der nicht adjustierten Simulationen	114
5.3.2. Ergebnisse bei adjustiertem a-priori Erwartungswert	129
5.3.3. Resümee der Simulationsstudie	132

IV. Diskussion

6. Zusammenfassung und Schlussfolgerungen	135
--	------------

Appendix

A. Gegenbeispiel Proportionalität von Fallzahl und Präzision	i
B. Anhänge zur Simulationsstudie	iii
B.1. Tabellen und Grafiken	iii
B.1.1. Ergebnisse der nicht adjustierten Simulationen	iii
B.1.2. Adjustierte Simulationsergebnisse	xxx
C. Programmsyntax	xxxiv
C.1. WinBUGS/OpenBUGS Code der Anwendungsbeispiele	xxxiv
C.1.1. Code für die Heparin-Studien	xxxiv
C.1.2. Code für die Metformin-Studien	xxxv

C.1.3. Code für die FeNO-studien	xxxv
C.2. R-Skript zur Berechnung von N_{prior}	xxxvii
C.3. R-Skript der Simulationsstudie	xli

Tabellenverzeichnis

Übersicht der Notationen.	11
4.1. Heparin-Studien: Ergebnisse der meta-analytisch prädiktiven Methode	83
4.2. Heparin-Studien: Poweranalyse	86
4.3. Heparin-Studien: Bestimmung von N_{prior}	89
4.4. Metformin-Studien: Ergebnisse der meta-analytisch prädiktiven Methode . . .	92
4.5. FeNO-Studien: Ergebnisse der meta-analytisch prädiktiven Methode	101
4.6. FeNO-Studien: Poweranalyse	103
4.7. FeNO-Studien: Bestimmung von N_{prior}	105
5.1. Übersicht der Simulationsszenarien und Einstellungen	110
5.2. Deskriptive Statistiken für die Power mit fixiertem μ_0 in der neuen Studie . .	129
B.1. Deskriptive Statistiken für N_{eff} : Szenario A, $OR = 0.5$	v
B.2. Deskriptive Statistiken für N_{eff} : Szenario B, $OR = 0.7$	vi
B.3. Deskriptive Statistiken für das Verhältnis von N_{eff} im getrimmten Modell zu N_{eff} im vollen Modell	vii
B.4. Deskriptive Statistiken für ω : Szenario A, $OR = 0.5$	x
B.5. Deskriptive Statistiken für ω : Szenario B, $OR = 0.7$	xi
B.6. Deskriptive Statistiken für $\mathbb{E}(\theta^* y)$: Szenario A, $OR = 0.5$	xiii
B.7. Deskriptive Statistiken für $\mathbb{E}(\theta^* y)$: Szenario B, $OR = 0.7$	xiv
B.8. Deskriptive Statistiken für $\mathbb{E}(\tau y)$: Szenario A, $OR = 0.5$	xvi
B.9. Deskriptive Statistiken für $\mathbb{E}(\tau y)$: Szenario B, $OR = 0.7$	xvii
B.10. Deskriptive Statistiken für $V_\tau^{1/2}$: Szenario A, $OR = 0.5$	xix
B.11. Deskriptive Statistiken für $V_\tau^{1/2}$: Szenario B, $OR = 0.7$	xx
B.12. Deskriptive Statistiken für die Power der neuen Studie: Szenario A, $OR = 0.5$	xxii
B.13. Deskriptive Statistiken für die Power der neuen Studie: Szenario B, $OR = 0.7$	xxiii
B.14. Deskriptive Statistiken für N_{prior} : Szenario A, $OR = 0.5$	xxviii
B.15. Deskriptive Statistiken für N_{prior} : Szenario B, $OR = 0.7$	xxix
B.16. Deskriptive Statistiken für N_{prior} adjustiert: Szenario A, $OR = 0.5$	xxx
B.17. Deskriptive Statistiken für N_{prior} adjustiert: Szenario B, $OR = 0.7$	xxxii

Abbildungsverzeichnis

2.1. Externe Evidenz	33
2.2. Schema eines hierarchischen Modells	36
3.1. Schema des vollen hierarchischen Modells.	58
3.2. Smooth-Scatterplots der α -Trimming-Methode	73
4.1. Forestplot der Heparin-Studien	77
4.2. MCMC-Diagnostiken für τ im Full-Modell der Heparin-Studien	80
4.3. MCMC-Diagnostiken für θ_{ped}^* im Full-Modell der Heparin-Studien	81
4.4. Forestplot der Metformin-Studien	90
4.5. Forestplot der FeNO-Studien	94
4.6. Dichten der Interstudienstandardabweichung τ für die FeNO-Studien	96
4.7. MCMC-Diagnostiken für τ im Full-Modell der FeNO-Studien	97
4.8. MCMC-Diagnostiken für θ_{ped}^* im Full-Modell der FeNO-Studien	98
5.1. Scatterplots: Vergleich zwischen N_{eff} im vollen und im getrimmten Modell in Szenario A bei 3 und 5 Studien	117
5.2. Scatterplots: Vergleich der bayesianische Power im vollen und im getrimmten Modell in Szenario A bei 3 und 5 Studien	123
5.3. Scatterplot: Vergleich der a-posteriori prädiktiven Erwartungswerte von θ^* und der bayesianischen Power in Szenario A bei 6 Studien	124
5.4. Zusammenhang zwischen $\mathbb{E}(\theta^* y)$, $V_\tau^{1/2}$, N_{eff} und bayesianischer Power	126
5.5. Scatterplots: Zusammenhang zwischen N_{prior} adjustiert und N_{eff}	131
B.1. Boxplots der Precision Weight Effective Sample Size N_{eff}	iv
B.2. Scatterplots: Vergleich zwischen N_{eff} im vollen und im getrimmten Modell	viii
B.3. Boxplots der Gewichtungsfaktoren ω	ix
B.4. Boxplots der a-posteriori prädiktiven Erwartungswerte von θ^*	xii
B.5. Boxplots der a-posteriori Erwartungswerte von τ	xv
B.6. Boxplots der a-posteriori prädiktiven Standardabweichung $V_\tau^{1/2}$ von θ^*	xviii
B.7. Boxplots der bayesianischen Power	xxi
B.8. Scatterplots: Vergleich der Power im vollen und getrimmten Modell	xxiv
B.9. Scatterplots: Vergleich von $\mathbb{E}(\theta^* y)$ und bayesianischer Power	xxv
B.10. Scatterplots: Zusammenhang zwischen N_{eff} und Power	xxvi
B.11. Boxplots der Prior Effective Sample Size N_{prior}	xxvii
B.12. Boxplots der adjustierten Prior Effective Sample Size N_{prior}	xxx
B.13. Scatterplots: Zusammenhang zwischen N_{prior} adjustiert und N_{eff}	xxxiii

Notationen

Bevor wir mit der eigentlichen Arbeit beginnen, werden wir einige wichtige Notationen und Abkürzungen auflisten, die in dieser Arbeit häufig verwendet werden. Betrachten wir gleichzeitig historische Studien, die aus zwei Kollektiven bestehen, so unterscheiden wir die Parameter für jedes Kollektiv durch Indizes. Eine Indizierung mit a bzw. ad bedeutet, dass sich der entsprechende Parameter auf ein Kollektiv erwachsener Patienten bezieht, während eine Indizierung mit k bzw. ped für einen Parameter in einem Kollektiv von Kindern steht. Weiter sei θ^* der zu schätzende Parameter in einer neuen Studie. In den klinischen Beispielen 4.1 und 4.3, sowie den simulierten Studienszenarien in Kapitel 5 wird dieser Parameter das prädiktive Log Odds Ratio sein. In den Metforminstudien in 4.2 ist θ^* die standardisierte Mittelwertveränderung in einer zukünftigen Studie. Die Begriffe Dichte und Verteilung verwenden wir synonym. Weitere wichtige Bezeichnungen sind in der folgenden Tabelle aufgelistet.

Notation	Bezeichnung
OR	Odds Ratio
LogOR	logarithmiertes Odds Ratio
p_T	Ereigniswahrscheinlichkeit der Behandlungsgruppe
p_C	Ereigniswahrscheinlichkeit der Kontrollgruppe
p_{TC}	gepoolte Ereigniswahrscheinlichkeit für Behandlungs- und Kontrollgruppe zusammen
τ^2	Varianz der Behandlungseffekte zwischen den historischen Studien (Interstudienvarianz)
τ	Standardabweichung der Behandlungseffekte zwischen den historischen Studien (Interstudienstandardabweichung)
ω	Verhältnis der a-posteriori prädiktiven Präzision der historischen Studien mit Wertverlust zur a-posteriori prädiktiven Präzision der historischen Studien ohne Wertverlust
ω_{ad}	ω für eine neue Erwachsenenstudie
ω_{ped}	ω für eine neue Kinderstudie

N_{eff}	Precision Weight Effective Sample Size
$N_{ad,eff}$	Precision Weight Effective Sample Size für eine neue Erwachsenenstudie
$N_{ped,eff}$	Precision Weight Effective Sample Size für eine neue Kinderstudie
N_{true}	wahre Precision Weight Effective Sample Size bei bekanntem τ
N_{prior}	Prior Effective Sample Size
$N_{ad,prior}$	Prior Effective Sample Size für eine neue Erwachsenenstudie
$N_{ped,prior}$	Prior Effective Sample Size für eine neue Kinderstudie
N_{hist}	Gesamtfallzahl der historischen Studien
μ_{shift}	systematischer Verschiebungsparameter des mittleren Behandlungseffekts zwischen Erwachsenen- und Kinderkollektiv
μ_0	fester Hyperparameter des mittleren Behandlungseffekts
τ_0	fester Hyperparameter der Standardabweichung des mittleren Behandlungseffekts
H	Anzahl der historischen Studien
h	$\in \{1, 2, \dots, H\}$
y_h	beobachteter Effekt in historischer Studie h
θ_h	wahrer studienspezifischer Effekt für historische Studie h
θ^*	vorhergesagter Behandlungseffekt in einer neuen Studie
A	Anzahl der historischen Studien an Erwachsenen
a	$\in \{1, 2, \dots, A\}$
y_a	beobachteter Effekt in Erwachsenenstudie a
θ_a	wahrer studienspezifischer Effekt für Erwachsenenstudie a
θ_{ad}^*	vorhergesagter Behandlungseffekt in einer neuen Erwachsenenstudie
$V_{ad,\tau,\mu_{shift}}$	Varianz von θ_{ad}^* bei variablem τ und μ_{shift}
$V_{ad,\tau\equiv 0,\mu_{shift_0}}$	Varianz von θ_{ad}^* mit $\tau \equiv 0$ und festem $\mu_{shift} = \mu_{shift_0}$
K	Anzahl der historischen Studien an Kindern
k	$\in \{A + 1, A + 2, \dots, A + K\}$

y_k	beobachteter Effekt in Kinderstudie k
θ_k	wahrer studienspezifischer Effekt für Kinderstudie k
θ_{ped}^*	vorhergesagter Behandlungseffekt in einer neuen Kinderstudie
$V_{ped,\tau,\mu_{shift}}$	Varianz von θ_{ped}^* bei variablem τ und μ_{shift}
$V_{ped,\tau\equiv 0,\mu_{shift_0}}$	Varianz von θ_{ped}^* mit $\tau \equiv 0$ und festem μ_{shift}
$p(\cdot)$	Verteilungs- bzw. Dichtefunktion
$p(\cdot \cdot)$	bedingte Verteilungs- bzw. Dichtefunktion
$N_\theta(\mu, \tau^2)$	bezeichnet die Dichtefunktion für $\theta \sim N(\mu, \tau^2)$
$\mathbb{P}(\cdot)$	Wahrscheinlichkeit für ein Ereignis, z. B. $\mathbb{P}(\theta < 0) = \int_{-\infty}^0 p(\theta) d\theta$

I Einleitung

1 Motivation

In dieser Arbeit untersuchen wir eine mathematische Methode, die es ermöglicht, Ergebnisse aus historischen klinischen Studien an Erwachsenen- und Kindern zu extrapolieren. Die historischen Daten sollen für eine prospektive Studie genutzt werden. Das Ziel einer solchen Extrapolation von Studienergebnissen besteht darin, Patienten bei der Rekrutierung einer zukünftigen Erwachsenen- bzw. Kinderstudie einzusparen. Einerseits können Patienten in der neuen Kontrollgruppe eingespart werden, wenn nur die Information der historischen Kontrollen genutzt wird. Andererseits ist es ebenfalls unter Berücksichtigung der äußeren Rahmenbedingungen und der jeweiligen zu untersuchenden Fragestellung möglich, die historischen Ergebnisse der Kontroll- und Behandlungsgruppe gleichzeitig zu nutzen. Dadurch kann in beiden Behandlungsarmen die benötigte Fallzahl reduziert werden. Wir betrachten in dieser Arbeit die simultane Extrapolation für beide Behandlungsgruppen. Zur Bewertung des Informationsgehaltes der historischen Studien verwenden wir einen meta-analytisch prädiktiven Bayes-Ansatz (Kap. 3). Das Ergebnis ist die a-posteriori prädiktive Verteilung (2.4) des studienspezifischen Behandlungseffekts einer neuen Studie. Im Rahmen eines hierarchischen Normal-Normal-Modells (Kap. 2.4) verwenden wir die a-posteriori prädiktive Verteilung als informative a-priori Verteilung einer neuen Studie. Diese informative Verteilung können wir in eine Anzahl virtueller Patienten übersetzen, die in einer spezifischen zukünftigen klinischen Studie eingespart werden können. Diese Anzahl bezeichnen wir als die Prior Effective Sample Size N_{prior} , also die Anzahl Patienten, die der a-priori Information entspricht (vgl. Kap. 3.4). Unter Anpassung eines hierarchischen Referenzmodells können wir einen Gewichtungsfaktor angeben, der das Ausmaß des Wertverlust in Form der Präzisionen der historischen Ergebnisse ausdrückt (vgl. Kap. 3.3). Die Herleitung des Gewichtungsfaktors ist objektiv und folgt asymptotischen Argumenten. Um die Wertminderung ebenfalls in Form einer Fallzahl auszudrücken, wird dieser Gewichtungsfaktor auch in eine virtuelle Anzahl Patienten, der Precision Weight Effective Sample Size, übertragen (vgl. 3.3.2). Diese Umrechnung ist jedoch intuitiv, da die Fallzahl nicht unbedingt im proportionalen Verhältnis zu Präzision stehen muss (vgl. Bspl. 3.1, A.1). Die Precision Weight Effective Sample Size werden wir in klinischen Anwendungsbeispielen und Simulationszenarien validieren.

Der entscheidende Parameter für das Ausmaß der Extrapolation der historischen Ergebnisse ist die Interstudienvarianz. Im Kontext der hierarchischen Modelle untersuchen wir, welche hyper a-priori Verteilungen für die Interstudienvarianz geeignet sind. In verschiedenen

klinischen Beispielen zeigen wir die Stärken und Schwächen dieser Methodik. Insbesondere betrachten wir die Situation, wenn nur wenige frühere Studien vorhanden sind. Wir wollen in dieser Arbeit vor allem Antworten auf die folgenden Fragen geben:

- Entspricht die Precision Weight Effective Sample Size der a-priori Information über den zukünftigen Behandlungseffekt einer zukünftigen Studie?
- Wird die bayesianische Power in einer neuen Studie unter Verwendung der Precision Weight Effective Sample Size eingehalten?
- Stimmt die Precision Weight Effective Sample Size mit der Prior Effective Sample Size überein?
- Wie viele historische Studien sind nötig, um den studienspezifischen Effekt einer zukünftigen Studie stabil und präzise vorherzusagen?
- Welche hyper a-priori Verteilung für die Interstudienvarianz ist bei geringer Anzahl an Vorstudien geeignet?

Die Fragen stehen im direkten Zusammenhang. Um diese zu beantworten, werden wir drei klinischen Anwendungsbeispiele sowie unterschiedliche, simulierte Studienszenarien betrachten (vgl. Kap. 4). Zusätzlich wenden wir eine neue Strategie, ein α -Trimming-Verfahren für die Markovketten, an, um die Stabilität der Parameterschätzungen zu beurteilen (vgl. Kap. 3.6).

1.1 Fallzahlproblematik

Heutzutage wird die Anzahl potentiell verfügbarer Patienten für klinische Studien immer kleiner. Dies hat unterschiedliche Gründe. Beispielsweise hat sich in den letzten Jahren besonders die medizinische Diagnostik mit Hilfe von Genexpressionsdaten, Zytologie oder bildgebenden Verfahren verbessert. Einzelne Krankheitsbilder können früher erkannt und immer genauer beschrieben werden. Dadurch nimmt die Anzahl der zu untersuchenden Subgruppen zu. Neue Therapiestrategien werden zunehmend speziell für diese Untergruppen entwickelt. Im Anschluss muss der Erfolg der neuen Behandlungskonzepte für die Subkollektive in klinischen Studien überprüft werden. Dadurch reduziert sich die Anzahl potentieller Studienteilnehmer drastisch. Ein anderes Gebiet mit kleinen Fallzahlen, stellen seltene Erkrankungen dar. In

diesem Bereich wird die Wirksamkeit von Arzneimitteln häufig nur deskriptiv beschrieben, da für prospektive Studien die notwendigen Studienteilnehmerzahlen nicht rekrutiert werden können. Nicht selten handelt es sich hierbei um einzelne Fallbeschreibungen.

In pädiatrischen Studien herrscht ebenfalls oft Patientenmangel. Häufig sind dies Studien aus dem Bereich der pädiatrischen Onkologie. Da die Inzidenzrate bei Kindern glücklicherweise sehr niedrig ist, stehen aber nur wenige potentielle Probanden zur Verfügung. Die Studiendauer erhöht sich dadurch so drastisch, dass die Finanzierung und Durchführung nicht mehr realisierbar werden. An Krebs erkrankte Kinder und Jugendliche werden meist im Rahmen von Therapiestudien behandelt. Patienten mit seltenen Tumoren, die nicht in Studien eingeschlossen werden, sind zum Beispiel in Registern wie dem *Seltene Tumoren-Register (STEP)*¹ erfasst.

In der klinischen Forschung ist der Status quo für gewöhnlich, dass die Wirksamkeit und Nebenwirkungen der Therapien bereits in einer Vielzahl klinischer Studien an Erwachsenen untersucht worden sind. Für den Einfluss auf Kinderpatienten liegen aber keine oder nur unzureichende Erkenntnisse vor. Zu einem speziellen Krankheitsbild gibt es dann viele historische Ergebnisse aus vorangegangenen klinischen Studien an Erwachsenen, während nur wenige Daten aus pädiatrischen Studien existieren. Allerdings lassen sich die Schlussfolgerungen aus den Analysen an Erwachsenen nicht ohne Weiteres direkt auf die Wirksamkeit bei Kindern übertragen. Kinder reagieren auf Medikamente und Therapien nicht genauso wie Erwachsene. Kinder sind keine kleinen Erwachsenen! Für konfirmatorische klinische Studien werden häufig hohe Fallzahlen benötigt, die allerdings allein durch herkömmliche Rekrutierungsmethoden nicht mehr erreicht werden können. Das Ziel der hier vorgestellten Methodik ist es, die Fallzahl in einer zukünftigen klinischen Studie zu reduzieren. Dies kann dadurch erreicht werden, dass Vorkenntnisse aus historischen Studien zur Planung einer neuen Studie genutzt werden. Die Studienplaner müssen hierfür sorgfältig überprüfen, ob die historischen Quellen die Voraussetzung für eine Extrapolation der Ergebnisse erfüllen (vgl. Kap. 1.2). Ferner muss in geeigneter Weise möglichst objektiv beurteilt werden, in welchem Umfang die Informationen generelle Aussagen über den Therapieeffekt in einer neuen Studie zulassen. Die Anwendung von Bayes-Verfahren bietet hierfür mehrere Modellierungsmöglichkeiten. Das Vorwissen fließt in Form einer a-priori Verteilung für den studienspezifischen Effekt in die Planung und Auswertung einer neuen Studie ein. Beispielsweise verwenden *Goodman et al.* [25] und *Greenhouse et al.* [28] Methoden aus der Bayes-Statistik, um auch das Vorwissen über

¹ www.kinderkrebsinfo.de/step

Erwachsene auf Kinder zu übertragen. Die Vorinformationen über die primäre Zielgröße in der Kontrollgruppe werden als informative a-priori Verteilung (vgl. Kapitel 2.1) modelliert und mit den Ergebnissen der neuen Studie über die Likelihood (vgl. Formel (2.2)) kombiniert. Die Informationen der Erwachsenenstudien werden herabgewichtet, um den Unterschieden zwischen den beiden Grundgesamtheiten, Kindern und Erwachsenen, gerecht zu werden [23, 25, 28, 50]. In *Goodman et al.* [25, S. 308] wird dieser Gewichtungsfaktor für den gepoolten Behandlungseffekt an Erwachsenen willkürlich auf 2 gesetzt. Die Varianz des gepoolten Effekts der früheren Erwachsenenstudien wird dadurch verdoppelt. Gleichzeitig wird die a-priori Verteilung der Kontrollgruppen der Erwachsenen auf 0 zentriert. Dieses Vorgehen ist allerdings suboptimal, da kein objektives Kriterium dafür spricht, genau diesen Gewichtungswert zu wählen.

Die Relevanz passende Designs für klinische Studien mit kleiner Fallzahl zu entwickeln, sowie der Nutzen von Bayes-Methoden sind mittlerweile auch von den internationalen Gesundheitsbehörden erkannt. So veröffentlichten die *European Medicines Agency (EMA)* und *U.S. Food and Drug Administration (FDA)* Richtlinien für die Anwendung von Bayes-Verfahren in klinischen Studien [67] und für die Durchführung klinischer Studien in kleinen Populationen [17]. Die Relevanz dieser Themen zeigt sich außerdem darin, dass spezielle Workshops wie beispielsweise *The Science of Small Clinical Trials (2011)* von der *FDA* durchgeführt wurden. In einzelnen Komitees wie dem *Paediatric Committee (PDCO)* der *EMA* wurden spezielle Arbeitsgruppen gegründet, die sich mit ebendiesen Anwendungen befassen.

1.2 Fallstricke historischer Daten

Historische Informationen unterliegen einer Vielzahl potentieller Fehlerquellen. Darum wird die Nutzung von historischen Kontrollgruppen kontrovers diskutiert, wie die folgenden Auszüge zeigen. In *Sacks et al.* [55] werden Studien mit randomisierten Kontrollen (RCTs) und Studien mit historischen Kontrollen (HCTs) verglichen, die zur gleichen Behandlung durchgeführt wurden. Dabei werden folgende Ergebnisse beobachtet:

„We found six therapies for which 50 RCTs and 56 HCTs were reported. Forty-four of 56 HCTs (79 percent) found the therapy better than the control regimen, but only 10 of 50 RCTs (20 percent) agreed.“ [Daraus wird der Schluss gezogen, dass [...]; Anm. d. V.]

„[...] *biases in patient selection may irretrievably weight the outcome of HCTs in favor of new therapies.*“

Doug Altman [3] beschreibt ebenfalls sehr kritisch die Zusammensetzung der historischen Kontrollen:

„*Problems will arise from the mixture of retrospective and prospective studies, and we can never satisfactorily eliminate possible biases due to other factors (apart from treatment) that may have changed over time.*“

Die Daten aus früheren Studien können, wie in den Quellen [3, 55] erwähnt, in mehrfacher Weise verzerrt sein. Beispielsweise können die Ergebnisse durch zeitliche Effekte beeinflusst werden. Demnach können sich die Diagnose- oder Therapiemöglichkeiten während der Zeit geändert haben, sodass die historischen Daten nicht mehr dem aktuellen Behandlungskonzept entsprechen. Andererseits können bei den unterschiedlichen Studien verschiedene Vergleichstherapien durchgeführt worden sein, sodass die Behandlungseffekte je nach Kontrollgruppe mehr oder weniger stark ausfallen können. Somit kann es in einigen Studien einen Vorteil der konventionellen Therapie geben, in anderen wiederum einen Vorteil der experimentellen Behandlung. Zudem können sich die äußeren Rahmenbedingungen wie Gesundheitspolitik, Compliance und Zusammensetzung der Kollektive verändert haben.

Einen großen Umbruch in der klinischen Forschung brachte die 12. AMG-Novelle im Jahr 2004¹. In dieser Gesetzesänderung wurden umfassende Neuregelungen zur Anwendung der guten klinischen Praxis eingeführt, die diese rechtsverbindlich machte (GCP-Verordnung (GCP-V) (§ 12, 42 AMG))². Die Neuregelungen in § 4 AMG und § 3 GCP-V betreffen beispielsweise die genaue Definition der Begriffe klinische Prüfung, Sponsor, Prüfer, Prüfplan und Prüfpräparat. Ein weiteres Resultat ist die Erhöhung der Transparenz klinischer Studien. Seit der 12. AMG-Novelle müssen, zusätzlich zum Zulassungsantrag eines Arzneimittels für Kinder und Jugendliche, Unterlagen über eine klinische Prüfung in diesem Patientenkollektiv eingereicht werden (BGBl. I S. 2535, S. 2037). Zuvor konnten die Ergebnisse klinischer Prüfungen an Erwachsenen einfach auf Kinder übertragen werden. Darum ist besondere Vorsicht bei der Kombination von Ergebnissen aus klinischen Studien geboten, die vor und nach der Einführung der 12. Arzneimittelgesetz-Novelle durchgeführt wurden. Hierbei dürfen wir

¹ Bundesgesetzblatt Jahrgang 2004 Teil I Nr. 41, ausgegeben zu Bonn am 5. August 2004. Zwölftes Gesetz zur Änderung des Arzneimittelgesetzes. Vom 30. Juli 2004.

² Gesetz über den Verkehr mit Arzneimitteln (Arzneimittelgesetz - AMG). GCP-Verordnung vom 9. August 2004 (BGBl. I S. 2081), die zuletzt durch Artikel 4 der Verordnung vom 3. November 2006 (BGBl. I S. 2523) geändert worden ist.

in der Regel weder von Struktur- noch Beobachtungsgleichheit der Studien ausgehen. Die Adjustierung bezüglich prognostischer Faktoren gestaltet sich demgemäß ebenfalls schwierig. Ebenso kann die Zusammensetzung der Kollektive der neuen Studie und der historischen Daten verschieden sein. Wenn historische Kontrollgruppen für eine neue Studie genutzt werden sollen, ist es wichtig die Studien nicht selektiv auszuwählen. Für den Fall, dass nur die Studien in die Analyse einfließen, die die neue Therapie favorisieren, kann das Outcome der neuen Studie stark beeinflusst werden. So würden wir gerade die Studien einschließen, die den gewünschten Behandlungsunterschied in der neuen Studie am schnellsten zeigen.

Um derartige systematische Fehler erst gar nicht zu begehen, sind daher kontrollierte klinische Studien State of the Art. Allerdings ist die Durchführung nicht immer möglich (vgl. Kap. 1.1). Wie wir bereits gesehen haben, gibt es aber auch Szenarien, in denen es notwendig ist auf historische Studienergebnisse zurückzugreifen. Zum Beispiel, wenn die Inzidenzrate der Erkrankung sehr niedrig ist und nicht genügend Patienten für eine neue zwei- oder mehrarmige Studie rekrutiert werden können.

Bei dem Gebrauch historischer Kontrollen bzw. Kontrollgruppen und Therapiegruppen sollten wir wie folgt vorgehen: Die erste zu klärende Frage ist, ob eine kontrollierte klinische Studie möglich ist. Ist dies nicht der Fall und sollen historische Quellen genutzt werden, so ist die Auswahl der zu berücksichtigenden historischen Studien mit Bedacht zu treffen. Wir müssen mögliche Confounder identifizieren und diesbezüglich adjustieren. Weiter müssen klare durchsichtige Regeln befolgt werden, die die Gleichheit der Kohorten gewährleisten. Dabei dürfen wir nicht nur spezielle Subgruppen, beispielsweise solche mit einem möglichst großen Therapieeffekt auswählen, sondern müssen alle Daten berücksichtigen. Weiter muss die Vollständigkeit der Daten gesichert sein. Eine Möglichkeit zur passenden Studienausswahl, stellen die Richtlinien der Cochrane-Library dar [27]. Alle Studien, die zu der geplanten Fragestellung bereits durchgeführt wurden, sollen eingeschlossen und die darauf folgende Selektion nachvollziehbar dokumentiert werden. Um weitere Fehlermöglichkeiten bei der Auswahl historischer Kontrollgruppen zu vermeiden, können wir außerdem auf die Guideline [19] der *EMA* für die Anwendung von historischen Kontrollen zurückgreifen. In dieser Arbeit gehen wir davon aus, dass die historischen Daten in den untersuchten Beispielen diese Kriterien in angemessener Weise erfüllen.

Die amerikanische Gesundheitsbehörde *FDA* weist in ihren Richtlinien [67, Kap. 4.6] darauf hin, dass es möglich ist, „*strength*“ von anderen Studien zu leihen. Als geeignete Methodik zur Bewertung wie viel Stärke von den früheren Studien geborgt werden kann, empfiehlt die *FDA* hierarchische Bayes-Modelle und gibt sogar ein Anwendungsbeispiel. Weiter weist die *FDA*

darauf hin, dass der Informationsgehalt in eine entsprechende Fallzahl umgerechnet werden kann:

„For device trials, the amount of strength borrowed can be translated into sample size, and the extent of borrowing depends on how closely results from the current study (the study of interest) reflect the results from the previous studies [...]“ [67, S. 25]

Wie die Stärke der historischen Studien in eine Fallzahl übersetzt werden kann, beschreibt die *FDA* allerdings nicht. An diesem Punkt setzen wir an und stellen in dieser Arbeit Möglichkeiten der Umrechnung vor. Im Anschluss analysieren wir diese in Situationen, in denen insbesondere die Verfügbarkeit historischer Informationen sehr eingeschränkt ist.

II Methoden

2 Grundlagen

Ein meta-analytisch prädiktiver Ansatz mit einer objektiven Bestimmung eines Gewichtungsfaktors wird in *Neuenschwander et al.* [50] vorgestellt. Die Informationsqualität der früheren Studien wird neutral bewertet. Durch die Anpassung von hierarchischen Modellen wird der Informationsgehalt der historischen Daten in Form von Wahrscheinlichkeitsverteilungen evaluiert. Die hierarchischen Modelle dienen dazu, bayesianische Metaanalysen der historischen Studien durchzuführen. Das allgemeine Vorgehen lässt sich komprimiert wie folgt darstellen: Als erster Schritt wird eine bayesianische Metaanalyse für die historischen Studien angepasst und durchgeführt. Ein Resultat ist die a-posteriori prädiktive Verteilung (2.4) des studienspezifischen Effekts θ^* einer zukünftigen Studie. In der neuen Erwachsenen- oder Kinderstudie wird die a-posteriori prädiktive Verteilung als a-priori Verteilung für den neuen Parameter benutzt. Die a-posteriori prädiktive Verteilung beinhaltet die Unsicherheit über θ^* in der neuen Studie und ist somit ein Maß für den Informationsgehalt der vorangegangenen Studien. Um die Qualität und den Umfang des Informationsgehaltes greifbarer darzustellen, wird ein Gewichtungsfaktor ω berechnet, der die Herabgewichtung der gepoolten Präzision beschreibt. Dieser Gewichtungsfaktor ist das Verhältnis der Präzision, die das aktuelle Wissen über den Studieneffekt widerspiegelt, und der fiktiven Präzision, falls die historischen Daten keine Information verlieren würde. Für die Berechnung wird ein zweites Modell, das Referenzmodell (Kap. 3.3.1), angepasst. Anschließend wird dieser Gewichtungsfaktor mit der Gesamtanzahl der historischen Patienten multipliziert. Die daraus resultierende Anzahl virtueller Patienten, die Precision Weight Effective Sample Size N_{eff} , stellt damit ein ungefähres Maß dar, wie viele Patienten das historische Vorwissen wert ist.

Ziel ist es die Fallzahl einer neuen Studie, die ohne Einbringung von Vorwissen zur Überprüfung der Hypothese notwendig ist, zu reduzieren. Die genaue Fallzahlersparnis durch die Anwendung des meta-analytisch prädiktiven Ansatzes, hängt aber immer von der konkret zu überprüfenden Hypothese ab. Sie wird unter anderem vom Signifikanzniveau und der Power beeinflusst. Die eingesparte Fallzahl, die Prior Effective Sample Size, werden wir exemplarisch für einige klinische Beispiele berechnen.

In dieser Arbeit stellen wir die methodischen Aspekte des meta-analytisch prädiktiven Ansatzes im Rahmen einer vereinfachten Modellannahme vor. Wir nehmen an, dass die zu schätzenden Behandlungseffekte bzw. Zielgrößen normalverteilt und die Parameter der ver-

schiedenen Studien austauschbar sind (vgl. Kap. 2.3.1). Das bedeutet, die Beobachtungen und die Parameter sind jeweils Realisationen von unabhängig identisch verteilten Zufallsvariablen.

2.1 Grundlagen der Bayes-Statistik

In diesem Kapitel geben wir einen groben Überblick über die Grundlagen der Bayes-Statistik. Jeder Leser, der sich bereits mit den Prinzipien der bayesianischen Statistik auskennt, kann dieses Kapitel zur Auffrischung seines Wissens nutzen oder bedenkenlos überspringen. Unter anderem stellen wir vor, was eine Likelihood, eine a-priori, a-posteriori und a-posteriori prädiktive Verteilung sind. Insbesondere betrachten wir die Eigenschaften, die sich bei Schätzungen von normalverteilten Parametern ergeben. Dabei werden wir genauer auf die Gestalt der einzelnen Dichtefunktionen in diesem Spezialfall eingehen und diese für ein hierarchisches Modell explizit berechnen.

Die nachfolgenden Abschnitte beziehen sich im Wesentlichen auf *Gelman et al.* [22, S. 6ff.]. Wir benutzen aus Gründen der Vergleichbarkeit analoge Bezeichnungskonventionen. Diese sind mathematisch etwas unpräzise formuliert, dafür aber anschaulich. Mit der Notation $A|B$ bezeichnen wir die bedingte Wahrscheinlichkeit von A gegeben B . Eine Verteilungsfunktion $\mathbb{R} \mapsto [0, 1]$ wird durch $p(\cdot)$ repräsentiert. $p(\cdot|\cdot)$ bezeichnet eine bedingte Wahrscheinlichkeitsdichte. Die Begriffe „Verteilung“ und „Dichte“ verwenden wir synonym wie in [22, S. 7]. Verschiedene Verteilungen innerhalb einer Gleichung werden ebenfalls durch $p(\cdot)$ dargestellt. Um Verwirrungen zu vermeiden, benutzen wir die Notation $\mathbb{P}(\cdot)$ als Wahrscheinlichkeit für ein Ereignis, beispielsweise $\mathbb{P}(\theta < 0) = \int_{-\infty}^0 p(\theta) d\theta$. Die Schreibweise $N_{\theta}(\mu, \tau^2)$ ist eine genauere Bezeichnung der Dichtefunktion der Normalverteilung von θ , wobei hier die Beziehung $\theta \sim N(\mu, \tau^2)$ gilt. Der zu schätzende Parameter bzw. Parametervektor sei θ bzw. $\theta = (\theta_1, \dots, \theta_n)$ und die beobachteten Daten seien y bzw. $y = (y_1, \dots, y_n)$.

In der klassischen frequentistischen Statistik gehen wir davon aus, dass die zugrunde liegenden wahren Parameter unbekannt aber fest sind. Die beobachteten Daten fassen wir als einzelne Realisation von allen möglichen Datensätzen auf. Punktschätzer und Konfidenzintervalle werden unter der Nullhypothese H_0 durch $p(y|H_0)$ bestimmt. Die Bayes-Statistik unterscheidet sich von der frequentistischen Sichtweise in der Annahme, dass lediglich die Daten fest und gegeben sind. Die Parameter sind unbekannt und werden wahrscheinlichkeitstheoretisch interpretiert. Die Bayes-Statistik zeichnet sich dadurch aus, die Unsicherheit über

den Parameter θ eines Modells durch eine Wahrscheinlichkeitsverteilung zu quantifizieren. Die Wahrscheinlichkeitsverteilung des Parameters θ ist gegeben durch $p(\theta)$. Sie repräsentiert das Vorwissen über den Modellparameter θ vor Beobachtung der Daten y . Sie wird als **a-priori Verteilung** bezeichnet. Ein Kritikpunkt an der Bayes-Methodik ist häufig die Auswahl der a-priori Verteilung (vgl. Kap. 2.5.1). Die Entscheidung für oder gegen eine bestimmte Verteilung basiert auf keinen festen Regeln, sodass häufig Willkür und Subjektivität bei der Wahl unterstellt werden [29, S. 328]. Die Aufgabe von $p(\theta)$ besteht darin, die Unsicherheit über θ vor Beobachtung der Daten wiederzugeben. Sie kann durch Experten spezifiziert oder von historischen Daten abgeleitet werden. Um θ im Vorfeld nicht zu sehr einzuschränken und eine große Unsicherheit über θ auszudrücken, werden in der Regel nicht- oder schwach-informative a-priori Verteilungen gewählt. Bei der Schätzung von normalverteilten Zielgrößen ähneln die bayesianischen Ergebnisse dann denen einer frequentistischen Analyse (siehe Kap. 2.2.1, Kap. 2.2.2, Kap. 2.4.2).

Die bedingte Verteilung $p(y|\theta)$, also die Verteilung der Daten y bei gegebenen Parametern θ , ist im Allgemeinen über explizite Modellannahmen gegeben. Diese Verteilung wird die **Likelihood** genannt. Die Parameter θ werden als bekannt angenommen, sodass wir unter Berücksichtigung der Parameter die Wahrscheinlichkeit angeben können, dass bestimmte Beobachtungen bzw. Realisationen eintreten.

Die Grundlage der Bayes-Statistik bildet die Bayes-Formel [7]. Damit gilt für die gemeinsame Verteilung $p(\theta, y)$ der beobachteten Daten und der Parameter

$$p(\theta, y) = p(\theta)p(y|\theta). \quad (2.1)$$

Die bedingte Wahrscheinlichkeitsverteilung $p(\theta|y)$ der Modell-Parameter θ gegeben den beobachteten Daten y wird **a-posteriori Verteilung** genannt. Das a-priori Wissen über θ wird durch die Daten aktualisiert. Mit Formel (2.1) lässt sich die a-posteriori Verteilung durch die a-priori Verteilung und die Likelihood ausdrücken

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y|\theta)}{\int p(\theta)p(y|\theta)d\theta}. \quad (2.2)$$

Das Integral in (2.2) ist allerdings nur in Ausnahmefällen von geschlossener Form und damit direkt analytisch berechenbar. Für die Bestimmung werden daher häufig Markov-Chain-Monte-Carlo (MCMC) Verfahren [vgl. 20, 31, 48] angewendet, auf die wir später noch näher eingehen. Für ein festes y können wir das Integral als Konstante $c \in (-\infty, \infty)$ auffassen,

da es dann nicht mehr von θ abhängt. Das Integral wird als Normierungskonstante bezeichnet. Es dient dazu die Fläche der a-posteriori Verteilung auf 1 zu normieren, sodass sie eine Wahrscheinlichkeitsverteilung ist. Daher können wir (2.2) umschreiben zu

$$p(\theta, y) = \frac{p(\theta)p(y|\theta)}{\int p(\theta)p(y|\theta)d\theta} = \frac{1}{c} \cdot p(\theta)p(y|\theta) \propto p(\theta)p(y|\theta). \quad (2.3)$$

Die a-posteriori Verteilung ist somit ein Gewicht aus der a-priori Verteilung $p(y)$, also dem Vorwissen über den Parameter θ vor Beobachtung der Daten, und der Likelihood $p(y|\theta)$, der Verteilung der beobachteten Daten bedingt auf θ . Die empirischen Daten fließen lediglich über die Likelihoodfunktion in die a-posteriori Analyse ein. Bayes-Modelle bestehen im Allgemeinen immer aus drei Schritten. Als erstes wird ein Wahrscheinlichkeitsmodell für die unbekannt Parameter spezifiziert, das a-priori Wissen über die Parameter beinhaltet. Im nächsten Schritt wird dieses Vorwissen über die unbekannt Parameter durch die Daten erneuert. Dies geschieht, indem das Wahrscheinlichkeitsmodell auf die beobachteten Daten bedingt wird. Als letztes werden die Modellanpassung und die Sensitivität der Ergebnisse bezüglich der unterstellten Annahmen überprüft.

Für die Vorhersage möglicher neuer Beobachtungen y^* bei zugrunde liegenden beobachteten Daten y , benötigen wir die **a-posteriori prädiktive Verteilung**. Sie enthält die Unsicherheit über die Vorhersage und ordnet möglichen neuen Beobachtungen eine Wahrscheinlichkeit zu

$$\begin{aligned} p(y^*|y) &= \int p(y^*, \theta|y)d\theta \\ &= \int p(y^*|\theta, y)p(\theta|y)d\theta \\ &= \int p(y^*|\theta)p(\theta|y)d\theta. \end{aligned} \quad (2.4)$$

In den Gleichungen ist die a-posteriori prädiktive Verteilung als Mittel über alle möglichen bedingten Vorhersagen über die a-posteriori Verteilung von θ dargestellt. Die letzte Gleichung gilt, da y und y^* bedingt unabhängig bei gegeben θ sind. Für die a-priori und a-posteriori Erwartungswerte bzw. die a-priori und a-posteriori Varianzen gelten folgende Beziehungen [vgl. 22, S. 37]

$$\mathbb{E}(\theta) = \mathbb{E}(\mathbb{E}(\theta|y)) \quad \text{und} \quad \text{Var}(\theta) = \mathbb{E}(\text{Var}(\theta|y)) + \text{Var}(\mathbb{E}(\theta|y)). \quad (2.5)$$

In Worten ausgedrückt heißt das, dass der a-priori Erwartungswert der Durchschnitt aller möglichen a-posteriori Erwartungswerte bezüglich der Verteilung der möglichen Beobachtungen ist. Die a-posteriori Varianz ist durchschnittlich kleiner als die a-priori Varianz. Der Umfang der Verkleinerung hängt von der Variation der a-posteriori Erwartungswerte über die Verteilung der Beobachtungen ab.

2.2 Normal-Modell

In diesem Abschnitt betrachten wir die konkrete Gestalt der einzelnen Verteilungen für die Schätzung des Mittelwerts einer Normalverteilung mit bekannter Varianz (Details siehe [45, Kap. 3.4; 22, Kap. 2.6]). Die Verteilungen lassen sich für diesen Spezialfall analytisch berechnen. Die Annahme normalverteilter Daten wird in klinischen Studien häufig getroffen. Der Zentrale Grenzwertsatz hilft diese Annahme zu rechtfertigen. Wir nehmen in dem Normal-Modellen an, dass sowohl die Parameter als auch die Daten jeweils normalverteilt sind. Die Gestalt der Verteilungen in diesem Modell ist für uns fundamental, da wir die Ergebnisse später für die approximative Bestimmung des Gewichtungsfaktors und der heuristischen Herleitung der Precision Weight Effective Sample Size benötigen (siehe Kap. 3).

2.2.1. Normal-Modell bei einer Beobachtungen

Betrachten wir als erstes den Fall, dass nur eine Beobachtung y vorliegt. Die Ergebnisse hierfür können wir schnell auf den geläufigeren Fall, eine Stichprobe mit mehreren Beobachtungen, verallgemeinern. Wir nehmen an, dass die Beobachtung y aus einer Normalverteilung mit Erwartungswert θ und *bekannter* Varianz σ^2 stammt, $y \sim N(\theta, \sigma^2)$. Die Likelihooddichte ist damit

$$p(y|\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-\theta)^2}{2\sigma^2}\right).$$

Wesentlich ist die Voraussetzung, dass σ^2 bekannt ist. Mit den festen und bekannten Hyperparametern μ_0 und τ_0 nehmen wir an, dass für den Parameter $\theta \sim N(\mu_0, \tau_0^2)$ gilt. Unter dieser Annahme erhalten wir durch Umformung die a-posteriori Dichte

$$p(\theta|y) \propto p(y|\theta) \cdot p(\theta)$$

$$\begin{aligned}
&= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-\theta)^2}{2\sigma^2}\right) \cdot \frac{1}{\sqrt{2\pi}\tau_0} \exp\left(-\frac{(\theta-\mu_0)^2}{2\tau_0^2}\right) \\
&= \frac{1}{2\pi\sigma\tau_0} \exp\left(-\frac{1}{2}\left(\frac{(y-\theta)^2}{\sigma^2} + \frac{(\theta-\mu_0)^2}{\tau_0^2}\right)\right),
\end{aligned}$$

wobei wir alle Variablen außer θ als konstant auffassen. Auf diese Weise ergibt sich

$$p(\theta|y) \propto \exp\left(-\frac{(\theta-\mu_1)^2}{2\tau_1^2}\right). \quad (2.6)$$

Also ist θ bezüglich der Beobachtung normalverteilt, $\theta|y \sim N(\mu_1, \tau_1^2)$, mit a-posteriori Erwartungswert μ_1 und Varianz τ_1^2 . Diese haben die Gestalt

$$\mu_1 = \frac{\frac{1}{\tau_0^2}\mu_0 + \frac{1}{\sigma^2}y}{\frac{1}{\tau_0^2} + \frac{1}{\sigma^2}} \quad \text{und} \quad \frac{1}{\tau_1^2} = \frac{1}{\tau_0^2} + \frac{1}{\sigma^2}. \quad (2.7)$$

Somit besteht die a-posteriori Präzision im Spezialfall normalverteilter Daten und normalverteilter a-priori Verteilung aus der Präzision der Daten $1/\sigma^2$ plus der Präzision der a-priori Verteilung $1/\tau_0^2$. Wir können den a-posteriori Erwartungswert μ_1 als gewichtetes Mittel des a-priori Erwartungswertes μ_0 und dem beobachteten Wert y interpretieren. Die Gewichte sind hierbei proportional zu den Präzisionen. Betrachten wir nun die Extremfälle, in denen der a-posteriori Mittelwert entweder dem a-priori Mittelwert oder dem Beobachtungswert entspricht. Es gilt

$$\begin{aligned}
\mu_1 &= \mu_0, \quad \text{falls } y = \mu_0 \quad \text{oder} \quad \tau_0^2 = 0 \\
\mu_1 &= y, \quad \text{falls } y = \mu_0 \quad \text{oder} \quad \sigma^2 = 0.
\end{aligned} \quad (2.8)$$

Falls die Varianz $\tau_0^2 = 0$ ist, ist die a-priori Verteilung eine Punktverteilung und μ_1 wird direkt durch μ_0 bestimmt. Die Daten haben dann keinen Einfluss auf den a-posteriori Erwartungswert. Der Hyperparameter μ_0 ist unendlich präzise und bestimmt damit die a-posteriori Verteilung. Im entgegengesetzten Fall, wenn $\sigma^2 = 0$ ist, sind die Daten bestimmend. Die a-posteriori Verteilung ist dann auf den Datenwert konzentriert. Diese extremen Fälle benötigen wir später für die Herleitung des Referenzmodells und der Bestimmung des Gewichtungsfaktors (vgl. Kap. 3.2, 3.3.1, 3.3). Die a-posteriori prädiktive Verteilung zur Vorhersage zukünftiger Beobachtungen erhalten wir durch Integration mit Gleichung (2.4)

$$p(y^*|y) = \int p(y^*|\theta)p(\theta|y)d\theta \propto \int \exp\left(-\frac{(y^*-\theta)^2}{2\sigma^2}\right) \exp\left(-\frac{(\theta-\mu_1)^2}{2\tau_1^2}\right) d\theta. \quad (2.9)$$

Die a-posteriori prädiktive Verteilung ist ebenfalls eine Normalverteilung. Die Parameter der Verteilung, der prädiktive Erwartungswert und die Varianz, lassen sich unter Benutzung der Ergebnisse der a-posteriori Verteilung, der Tatsache, dass $\mathbb{E}(y^*|\theta) = \theta$ und $\text{Var}(y^*|\theta) = \sigma^2$ gilt, sowie der Gleichungen (2.5) bestimmen. Somit gilt

$$\mathbb{E}(y^*|y) = \mathbb{E}(\mathbb{E}(y^*|\theta, y)|y) = \mathbb{E}(\theta|y) = \mu_1, \quad (2.10)$$

$$\begin{aligned} \text{Var}(y^*|y) &= \mathbb{E}(\text{Var}(y^*|\theta, y)|y) + \text{Var}(\mathbb{E}(y^*|\theta, y)|y) \\ &= \mathbb{E}(\sigma^2|y) + \text{Var}(\theta|y) \\ &= \sigma^2 + \tau_1^2. \end{aligned} \quad (2.11)$$

Der Erwartungswert der a-posteriori prädiktiven Verteilung ist der gleiche wie der Erwartungswert der a-posteriori Verteilung. Die Varianz besteht allerdings aus zwei Komponenten, zum einen aus der Varianz des Modells σ^2 und zum anderen aus der a-posteriori Varianz τ_1 , die die Unsicherheit über θ beinhaltet.

2.2.2. Normal-Modell mit mehreren Beobachtungen

Wir betrachten nun die Erweiterung des obigen Modells für mehrere Beobachtungen. Die einzelnen Beobachtungen $y = (y_1, \dots, y_n)$ seien Realisierungen unabhängiger identisch normalverteilter Zufallsvariablen. Die Beobachtungen können zum Beispiel die Therapieergebnisse von einzelnen Patienten sein. Unter der Voraussetzung *bekannter* Varianz σ^2 ergibt sich nach Umformung [vgl. 22, Kap. 2.6] für die a-posteriori Dichte

$$\begin{aligned} p(\theta|y) &\propto p(\theta)p(y|\theta) \\ &\stackrel{u.i.v.}{=} p(\theta) \prod_{h=1}^n p(y_h|\theta) \\ &\propto \exp\left(-\frac{1}{2\tau_0^2}(\theta - \mu_0)^2\right) \prod_{h=1}^n \exp\left(-\frac{1}{2\sigma^2}(y_h - \theta)^2\right) \\ &\propto \exp\left(-\frac{1}{2}\left(\frac{1}{\tau_0^2}(\theta - \mu_0)^2 + \frac{1}{\sigma^2}\sum_{h=1}^n (y_h - \theta)^2\right)\right), \end{aligned} \quad (2.12)$$

mit a-priori Verteilung $\theta \sim N(\mu_0, \tau_0)$, festen Hyperparametern μ_0 und τ_0^2 , sowie der Standardabweichung σ^2 . Die a-posteriori Verteilung hängt von y nur durch den Stichprobenmittelwert $\bar{y} = \frac{1}{n}\sum_h y_h$ ab. Der Mittelwert ist in diesem Modell eine suffiziente Statistik (vgl. [1, S. 34ff.,

Def. 7.2, Kor. 8.5, Kor. 8.6]). Da damit $\bar{y}|\theta, \sigma^2 \sim N(\theta, \sigma^2/n)$ gilt, können wir \bar{y} wie eine einzelne Beobachtung interpretieren und somit die Ergebnisse für das Normal-Modell bei einer Beobachtung anwenden. Es folgt aus den Beziehungen (2.6) und (2.7) unmittelbar

$$p(\theta|y_1 \dots y_n) = p(\theta|\bar{y}) = N_\theta(\mu_n, \tau_n^2)$$

mit

$$\mu_n = \mathbb{E}(\theta|y_1 \dots y_n) = \frac{\tau^2}{\frac{\sigma^2}{n} + \tau_0^2} \bar{y} + \frac{\frac{\sigma^2}{n}}{\frac{\sigma^2}{n} + \tau_0^2} \mu_0 = \frac{\frac{1}{\tau_0^2} \mu_0 + \frac{n}{\sigma^2} \bar{y}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}} \quad (2.13)$$

und

$$\tau_n^2 = \frac{\sigma^2 \tau_0^2}{\sigma^2 + \tau_0^2} \quad \text{bzw.} \quad \frac{1}{\tau_n^2} = \frac{1}{\tau_0^2} + \frac{n}{\sigma^2}. \quad (2.14)$$

Durch Gleichung (2.13) wird der gleichbedeutende Einfluss der a-priori Präzision $1/\tau_0^2$ und der empirischen Präzision der Daten deutlich. Ist die Anzahl der Beobachtungen n groß, wird die a-posteriori Verteilung durch den Stichprobenmittelwert \bar{y} und die Stichprobenvarianz σ^2 bestimmt. Falls wir die a-priori Variation bei festem n gegen unendlich betrachten ($\tau_0 \rightarrow \infty$), bzw. falls wir bei festem τ_0^2 die Anzahl der Beobachtungen gegen unendlich laufen lassen ($n \rightarrow \infty$), so gilt asymptotisch die Approximation

$$p(\theta|y) \stackrel{\text{appr.}}{\approx} N_\theta\left(\bar{y}, \frac{\sigma^2}{n}\right). \quad (2.15)$$

Eine a-priori Verteilung mit großer Varianz τ^2 , also kleiner Präzision, drückt geringes Vorwissen aus und lässt einen großen Wertebereich für θ zu. Damit werden die posteriori Werte von θ durch die Likelihood bestimmt, während die a-priori Verteilung keinen Einfluss hat. Vergleichen wir die Ergebnisse in diesem Spezialfall mit denen in der klassischen Statistik, in der die Daten als Realisationen der unbekanntem und festen Parameter aufgefasst werden, so gilt unter den obigen Annahmen $\bar{y}|\theta, \sigma^2 \sim N(\theta, \sigma^2/n)$. In diesem Spezialfall führen darum die bayesianische und die frequentistische Herangehensweise zu den gleichen Verteilungen. In diesem Modell liefert der Bayes-Ansatz unter Verwendung nicht-informativer Prior somit dieselben Ergebnisse wie die klassische Statistik. Durch $\bar{y} \pm z_{0.975} \frac{\sigma}{\sqrt{n}}$ ist daher in beiden Fällen ein gültiges 95%-Konfidenzintervall bzw. 95%-Credible-Intervall für θ gegeben, wobei $z_{0.975}$ das 97.5%-Quantil der Standardnormalverteilung bezeichnet.¹

¹ In der Bayes-Statistik wird bei Intervallschätzungen in der Regel von Credible-Intervallen gesprochen. Diese sind im Allgemeinen nicht identisch mit frequentistischen Konfidenzintervallen. Im obigen Spezialfall stimmen jedoch beide überein. Da wir in dieser Arbeit die Credible-Intervalle approximativ unter der Normalverteilungsannahme berechnen, verwenden wir im Folgenden die Bezeichnung Konfidenzintervall.

2.3 Zusammenhänge von Studieneffekten

Die Annahme, wie die einzelnen Beobachtungen oder die Studieneffekte zusammenhängen, ist essentiell für die meta-analytisch prädiktive Methode. Wir beschränken uns in dieser Arbeit auf einen austauschbaren Zusammenhang, was heißt, dass die Parameter unabhängige Ziehungen aus einer gemeinsamen Wahrscheinlichkeitsverteilung sind. Trotzdem stellen wir nachfolgend eine Auswahl weiterer denkbarer und ebenfalls modellierbarer Zusammenhänge dar. Teils wird dadurch das angepasste Modell sehr komplex, teils stark vereinfacht. Die Annahme über die Art des Zusammenhangs sollte immer der jeweiligen Anwendungssituation entsprechen. Wenn wir später zeigen, wie die Daten von zwei Kollektiven kombiniert und die historischen Resultate gewichtet werden können, werden wir Ähnlichkeiten zu den anderen Beziehungsmodellen im nächsten Abschnitt 2.3.1 erkennen.

2.3.1. Modellannahmen und Evidenz

Der Zusammenhang zwischen den einzelnen Parametern der historischen Studien $\theta_1, \dots, \theta_H$ und dem studienspezifischen Parameter der zukünftigen Studie θ^* wird durch ein Modell beschrieben. Dieses Modell sei durch $\theta_1, \dots, \theta_H, \theta^* | \eta$ gegeben. Ohne Beschränkung der Allgemeinheit betrachten wir im Folgenden ein parametrisches Modell mit Modellparameter η . Die nachstehenden Relationen (a)-(f) werden in *Neuenschwander et al.* [50, S. 7], *Spiegelhalter et al.* [62, Kap. 3.16, S. 91f.; Kap. 5.4, S. 148f.], sowie in *Gelman et al.* [22, S. 121f.] beschrieben. In Abbildung 2.1 sind die Zusammenhangsmodelle grafisch dargestellt.

- (a) *Irrelevant*: Die historischen Ergebnisse enthalten keine Information über den Parameter θ^* der neuen Studie. Die a-priori Verteilung für die neue Studie wird unabhängig von den historischen Daten gewählt und das Vorwissen komplett ignoriert.
- (b) *Austauschbar*: Alle Studien sind *ähnlich*. Dies bedeutet, die Parameter der alten und neuen Studien $\theta_1, \dots, \theta_H, \theta^*$ stammen aus einer gemeinsamen Verteilung, beispielsweise aus einer Normalverteilung: $\theta_1, \dots, \theta_H, \theta^* \sim N(\mu, \tau^2)$. Dieser Fall entspricht der typischen Annahme in Metaanalysen, allerdings nun in einem bayesianischen Kontext. Für diesen Zusammenhang können wir relativ einfach hierarchische Modellierungstechniken anpassen. Aus dem Blickwinkel eines Frequentisten entspricht die Modellierung dieser Beziehung einer Random-Effects-Metaanalyse. Jedoch muss hier nicht vorausgesetzt

werden, dass die Studien wiederum aus einer Superpopulation stammen. Eine notwendige Annahme dagegen ist, dass die Studien sich nicht systematisch unterscheiden.

- (c) *Möglicherweise verfälscht*: Die historischen Studien sind verfälscht, entweder durch internen Bias, der durch mangelnde Qualität der Studie entstanden sein kann, oder durch externen Bias, wie zum Beispiel Ungenauigkeiten im Zielparameter, oder durch Kombination beider Fehlermöglichkeiten. Die einzelnen Parameter $\theta_1, \dots, \theta_H$ stellen Funktionen von θ^* dar. Üblicherweise wird angenommen, dass ein Verschiebungsparameter existiert, sodass $\theta_h = \theta^* + \delta_h$ mit $h \in \{1, \dots, H\}$.
- (d) *Gleich aber herabgewichtet*: Die Parameter der alten Studien sind unverfälscht und gleich dem Parameter der neuen Studie. Allerdings wird die Präzision des historischen Effektes vergrößert, wodurch der Informationsgehalt der früheren Studien herabgewichtet wird. Dieses Prinzip ist der „*power prior*“ Ansatz, der in [11, 33] beschrieben wird. Anwendungen in klinischen Studien finden sich in *Goodman et al.* [25] und *Greenhouse et al.* [28]. In [25, 28] wurde aber die Wahl des Gewichtungsfaktors nicht auf Grundlage objektiver Statistikverfahren getroffen, sondern subjektiv festgelegt. Mathematisch ausgedrückt bedeutet dieser Zusammenhang, dass einerseits $\theta_h = \theta^*$ für alle $h \in \{1, \dots, H\}$ gilt und dass andererseits die in der zugehörigen Likelihood $p(y_h|\theta_h)$ enthaltene Information um einen Faktor $\alpha \in [0, 1]$ reduziert wird. Hierzu betrachten wir das folgende Beispiel.

Beispiel 2.1: Die informative a-priori Verteilung für den Parameter θ^* der neuen Studie gegeben die historischen Studienergebnisse lautet unter der Annahme normalverteilter historischer Likelihoods

$$\theta^* | y_1, \dots, y_H \sim N \left(\frac{\sum_h y_h \frac{1}{\sigma_h^2}}{\sum_h \frac{1}{\sigma_h^2}}, \frac{1}{\alpha \sum_h \frac{1}{\sigma_h^2}} \right). \quad (2.16)$$

Für $\alpha = 0$ verlieren die historischen Ergebnisse vollständig an Bedeutung. Im Gegensatz dazu wird für $\alpha = 1$ der Einfluss der Studien nicht reduziert und die enthaltenen Informationen werden in vollem Umfang und Gewicht genutzt. Gibt es nur eine frühere Studie, reduziert sich (2.16) zu

$$\theta^* | y_h \sim N \left(y_h, \frac{\sigma_h^2}{\alpha} \right).$$

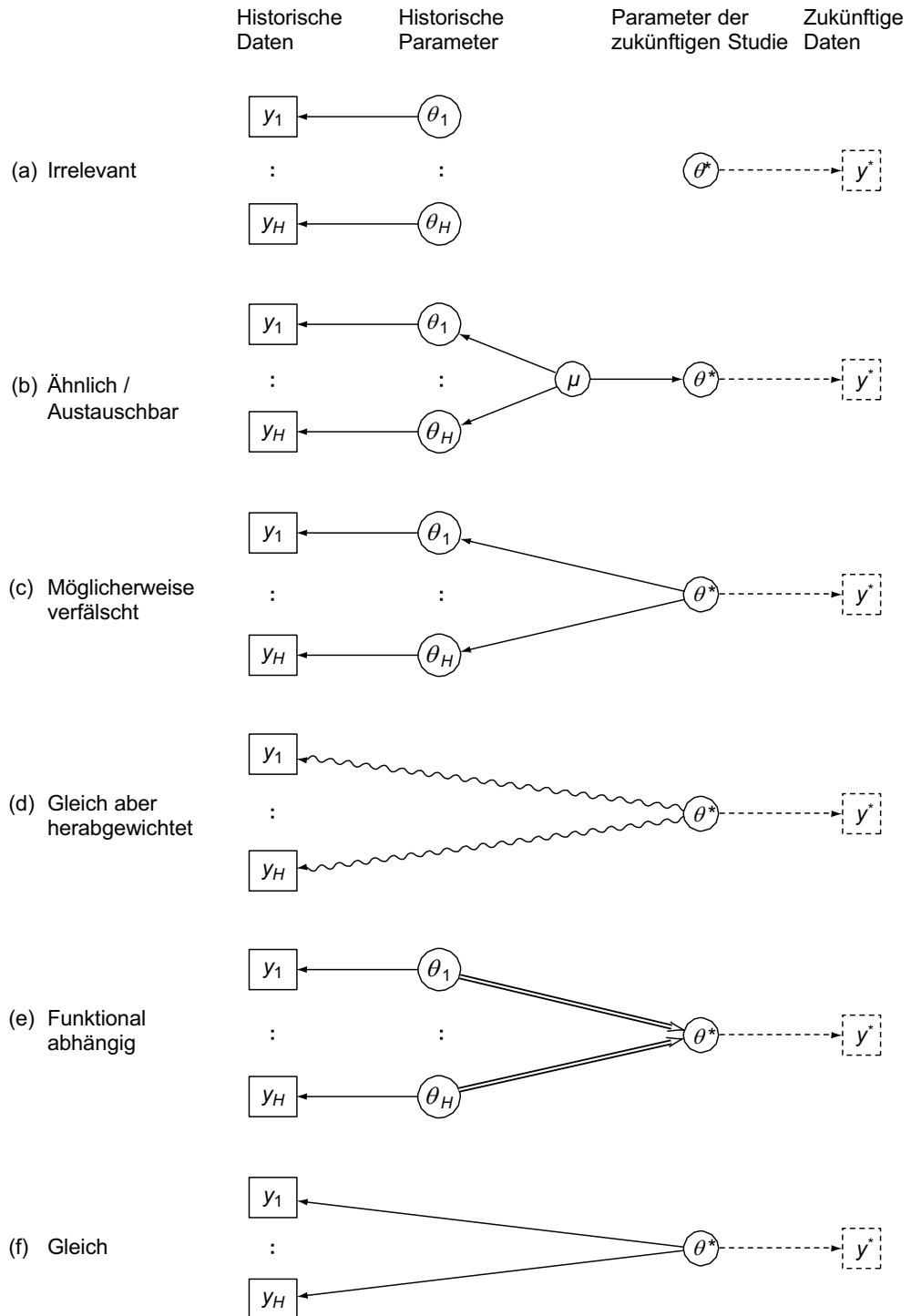


Abb. 2.1: Externe Evidenz.

Unterschiedliche Annahmen über den Zusammenhang zwischen den historischen Daten und den Parametern der neuen Studie. Einzelne Pfeile repräsentieren Verteilungen, doppelte Pfeile logische Funktionen und geschlängelte Pfeile Herabgewichtungen.

Quelle: vgl. Spiegelhalter et al. [62, S. 149].

- (e) *Funktional abhängig*: Die Zielgröße der neuen Studie ist eine logische Funktion bestehend aus den Parametern der historischen Studien.

Beispiel 2.2: Angenommen θ_1 sei der Behandlungseffekt einer Studie, die nur an Erwachsenen durchgeführt wurde, und θ_2 der Effekt aus einer Studie, die nur Kinder beinhaltet. Ein möglicher Behandlungseffekt einer zukünftigen Kinder- und Erwachsenenstudie mit einem Anteil von p Erwachsenen könnte sich durch den Zusammenhang

$$\theta^* = p\theta_1 + (1-p)\theta_2$$

beschreiben lassen. Die a-priori Verteilung für θ^* in der neuen Studie wird somit aus der gemeinsamen Evidenz der Parameter θ_1 und θ_2 ermittelt.

- (f) *Gleich*: In dieser Situation nehmen wir an, dass alle studienspezifischen Parameter der vergangenen Studien identisch sind. Dies setzt die strenge Annahme voraus, dass die individuellen Patienten bzw. Beobachtungen aller Studien austauschbar sind. Das bedeutet, jeder Patient könnte mit dem gleichen Ergebnis in einer anderen Studie behandelt worden sein. Diese Voraussetzung ist jedoch in den wenigsten Fällen gerechtfertigt. Die Studien unterscheiden sich meist in ihren Rahmenbedingungen. Zum Beispiel können sich die Studien aus multizentrischen und Single-Center-Studien oder aus prospektiven und retrospektiven Studien zusammensetzen, oder die räumliche Zusammensetzung der Studienpopulationen kann divergieren (vgl. Kap. 1.2).

Beispiel 2.3: Im normalverteilten Modell bedeutet diese Zusammenhangsannahme, dass die ursprünglichen Daten vollständig gepoolt werden können. Es gilt $\theta_1 = \theta_2 = \dots = \theta_H = \theta^*$. Das Vorwissen wird in voller Gewichtung als a-priori Verteilung eingebracht

$$\theta^* | y_1, \dots, y_H \sim N \left(\frac{\sum_h y_h \frac{1}{\sigma_h^2}}{\sum_h \frac{1}{\sigma_h^2}}, \frac{1}{\sum_h \frac{1}{\sigma_h^2}} \right). \quad (2.17)$$

Im Fall einer einzigen historischen Studie reduziert sich (2.17) zu

$$\theta^* | y_h \sim N \left(y_h, \sigma_h^2 \right).$$

In dieser Arbeit betrachten wir ohne Einschränkung den Ansatz eines *ähnlichen* Zusammenhangs. Der meta-analytisch prädiktive Ansatz wurde in [50, S. 7; 24, 30] ebenfalls unter derselben Annahme vorgestellt und weiterentwickelt. Ein austauschbarer Zusammenhang ist

keine allzu starke Voraussetzung, sondern die übliche Annahme, die oft in klassischen und bayesianischen Metaanalysen unterstellt wird.

2.4 Hierarchische Modelle

In vielen statistischen Anwendungen werden gleichzeitig mehrere Parameter betrachtet, die in irgendeiner Weise voneinander abhängen. Mit hierarchischen Modellen können diese multi-parametrischen Zusammenhänge geeignet modelliert werden. Der folgende Abschnitt bezieht sich auf [22, Kap. 5]. Die Formeln und Herleitungen können dort nachgeschlagen werden.

Hierarchische Modelle werden häufig für Metaanalysen verwendet. Die Grundstruktur eines hierarchischen Modells ist in Abbildung 2.2 dargestellt. Im Gegensatz zu den zuvor vorgestellten Normal-Modellen (Kap. 2.2.1, Kap. 2.2.2) gibt es nicht nur einen gemeinsamen Parameter, sondern mehrere Gruppen von Parametern. Auf der untersten Ebene stehen die tatsächlichen Beobachtungen y_{ih} , wobei $h = 1, \dots, H$ die Gruppen und $i = 1, \dots, n_h$ die Beobachtungen innerhalb der Gruppe indiziert. Die Beobachtungen können zum Beispiel die Patienten in einer klinischen Studie und die Gruppen beispielsweise die separaten Studien sein. Auf der nächsthöheren Ebene befinden sich die Gruppenparameter, die hier die wahren mittleren Behandlungseffekte θ_h für jede einzelne Studie darstellen. Mit $\theta := (\theta_1, \dots, \theta_H)$ bezeichnen wir den Vektor der Gruppenparameter. Die Variabilität unter den Gruppeneffekten θ_h können wir modellieren, indem wir sie als Stichprobe aus einer gemeinsamen Verteilung betrachten. Auf der obersten Ebene steht der Vektor ϕ der Hyperparameter. In unseren klinischen Anwendungsbeispielen (Kap. 4) besteht dieser Vektor zum Beispiel aus $\phi = (\mu, \mu_{shift}, \tau)$. Dort ist μ der gemeinsame wahre mittlere Behandlungseffekt über alle Erwachsenenstudien, μ_{shift} der systematische Verschiebungsparameter zwischen dem wahren gepoolten Behandlungseffekt bei Erwachsenen μ und dem bei Kindern $\mu + \mu_{shift}$, und τ ist die Standardabweichung der gepoolten Therapieeffekte, die wir in beiden Kollektiven als gleich annehmen. Eine detaillierte Beschreibung folgt in Kapitel 3. Das hierarchische Modell lässt sich damit so interpretieren, dass der Hyperparameter μ der wahre gepoolte Studieneffekt eines Kollektivs ist, um den die einzelnen wahren Studieneffekte θ_h streuen, um die wiederum die einzelnen Beobachtungen y_{ih} in jeder Studie streuen. Die Hyperparameter sind in der Regel nicht fest, sondern besitzen ebenfalls Verteilungen. Im Allgemeinen werden für die Hyperparameter nicht-informative a-priori Verteilungen gewählt, die den Wertebereich der Parameter nicht einschränken [22,

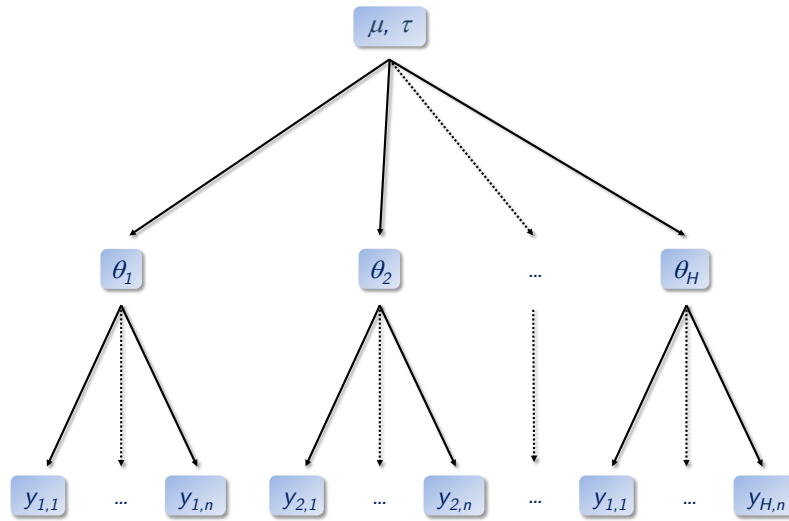


Abb. 2.2: Schema eines hierarchischen Modells.

Die Pfeile repräsentieren die Verbindungen zwischen den Knoten über Wahrscheinlichkeitsverteilungen.

S. 124f.; 34, S. 4; 67, S. 23]. In der einfachsten Situation gehen wir davon aus, dass alle θ_h austauschbar sind (vgl. Kap. 2.3.1b). Die Verteilung $p(\theta_1, \dots, \theta_H)$ ist somit invariant unter Permutationen von $1, \dots, H$ und die θ_h sind unabhängig identisch verteilt. Das bedeutet, die gemeinsame Verteilung der Parameter bedingt auf die Hyperparameter ist eine Produktverteilung

$$p(\theta|\phi) = p(\theta_1, \dots, \theta_H|\phi) \stackrel{u.i.v.}{=} \prod_h^H p(\theta_h|\phi). \quad (2.18)$$

Da ϕ in der Regel unbekannt ist, erhalten wir die gemeinsame Randverteilung der Gruppenparameter durch Integration über alle Kombinationen von ϕ

$$p(\theta) = p(\theta_1, \dots, \theta_H) = \int \prod_h^H p(\theta_h|\phi) p(\phi) d\phi. \quad (2.19)$$

Der Bayes-Ansatz besteht darin, die gemeinsame a-posteriori Verteilung aller unbekannt Parameter und Hyperparameter zu berechnen. Anschließend integrieren wir über die Nuisance-Parameter und erhalten die Randverteilung der uns interessierenden Parameter. Ist eine direkte Berechnung der gemeinsamen Verteilung aller Parameter nicht möglich, können wir auf Markov-Chain-Monte-Carlo (MCMC) Verfahren zurückgreifen. Mit dieser Technik generieren wir bedingte Stichproben aus der gemeinsamen a-posteriori Verteilung. Allerdings können wir die bedingten Verteilungen und die Randverteilungen in einigen Fällen auch analytisch herleiten. Das allgemeine Vorgehen zur analytischen Bestimmung der gemeinsamen a-posteriori Verteilung gliedert sich dabei in die folgenden Schritte:

1. Zerlege die gemeinsame a-posteriori Verteilung $p(\theta, \phi|y)$ in das nicht-normalisierte Produkt, bestehend aus der hyper a-priori Verteilung $p(\phi)$, der Verteilung der Gruppenparameter $p(\theta|\phi)$ und der Likelihood $p(y|\theta)$

$$p(\theta, \phi|y) \propto p(\phi, \theta)p(y|\phi, \theta) = p(\phi, \theta)p(y|\theta) = p(\phi)p(\theta|\phi)p(y|\theta). \quad (2.20)$$

Die erste Gleichung gilt, da die Hyperparameter ϕ die Daten y nur indirekt durch die Parameter θ beeinflussen.

2. Bestimme die bedingte a-posteriori Verteilung $p(\theta|\phi, y)$. Für eine feste Beobachtung y ist diese Verteilung eine Funktion in ϕ .
3. Schätze ϕ mit der Bayes-Formel, das heißt, bestimme die a-posteriori Randverteilung $p(\phi|y)$.

Wir sind insbesondere an der Verteilung $p(\theta^*|y)$ zur Vorhersage des Parameters θ^* einer neuen Studie interessiert. Dafür benötigen wir zunächst die gemeinsame a-posteriori Verteilung $p(\theta, \phi|y)$. Dabei folgen wir den Schritten 1.-3. und ziehen dann neue Parameter aus der zugehörigen a-posteriori Verteilung $p(\theta|y)$. Die konkrete Gestalt der bedingten Verteilungen ist nur selten bekannt, sodass sich die gemeinsame a-posteriori Verteilung nur in Ausnahmefällen analytisch berechnen lässt. Aus diesem Grund verwenden wir MCMC-Verfahren, die dieses Problem geeignet umgehen.

2.4.1. MCMC-Verfahren

Der Metropolis-Hastings Algorithmus bzw. der Gibbs-Sampler sind die populärsten Updating-Algorithmen für Markovketten [31, 48]. Sie sind Verfahren zur Erzeugung von Markovketten mit einer gewünschten stationären Verteilung. Gibbs-Sampling eignet sich vor allem dann, wenn die gemeinsame Verteilung $p(\theta, \phi|y)$ eines Zufallsvektors (θ, ϕ) unbekannt ist, jedoch die bedingten Verteilungen der einzelnen Komponenten bekannt sind. Das Grundprinzip besteht darin, wiederholend Werte für die einzelnen Komponenten gemäß ihrer bedingten Verteilung zu erzeugen, wobei diese Ziehungen von den vorherigen Werten abhängen. Die Werte aller anderen Komponenten bleiben im einzelnen Iterationsschritt unverändert. Aus der entstehenden Folge von Stichprobenvektoren lässt sich eine Markovkette herleiten. Die stationäre Verteilung dieser Markovkette ist gerade die gesuchte gemeinsame Verteilung aller Parameter.

2.4.2. Normal-Normal-Modell

In diesem Modell brauchen wir nicht auf MCMC-Verfahren zurückgreifen, sondern können die bedingten Verteilungen in Formel (2.20) analytisch bestimmen. Exemplarisch wollen wir dies einmal im Detail durchexerzieren.¹

In diesem Spezialfall eines hierarchischen Modells wollen wir normalverteilte Mittelwerte schätzen. Das Normal-Normal-Modell ist bekannt aus der klassischen Metaanalyse [51, S. 332f.; 69]. Wir können dieses Modell beispielsweise als eine bayesianische Metaanalyse mit nicht-informativer Gleichverteilung innerhalb eines Patientenkollektivs interpretieren. Hierbei handelt es sich um einen Sonderfall des später in Kapitel 3.2 vorgestellten vollen Modells. Die Hierarchiestruktur ist:

Normal-Normal-Modell:

$$y_{ih}|\theta_h \stackrel{u.i.v.}{\sim} N(\theta_h, \sigma^2) \quad \text{mit } i = 1, \dots, n_h, h = 1, \dots, H \quad (2.22)$$

$$\theta_h|\mu, \tau^2 \stackrel{u.i.v.}{\sim} N(\mu, \tau^2) \quad (2.23)$$

$$p(\mu, \tau) = p(\mu|\tau)p(\tau) \propto p(\tau) \propto \mathbb{1}_{\{\tau>0\}}. \quad (2.24)$$

Die Beobachtungen sind unabhängig identisch normalverteilt. Ebenso sind die Gruppenparameter θ_h unabhängige Ziehungen aus einer gemeinsamen Normalverteilung $N(\mu, \tau^2)$ mit den Hyperparametern μ und τ . Überdies nehmen an, dass die Stichprobenvarianz σ^2 bekannt ist. Auf der Hyperparameterebene (2.24) setzen wir voraus, dass die Verteilung für μ bei gegebener Varianz τ^2 eine nicht-informative Gleichverteilung $\mathbb{1}_{(-\infty, \infty)}$ ist. Das Integral über diese hyper a-priori Dichte ist uneigentlich. Die Verteilung ist improper. Improper bedeutet, dass das Integral über die normalisierte Dichte nicht 1 ergibt und proper, dass dieses gleich 1 ist. Wir lassen durch diese Verteilung alle möglichen Werte für $\mu \in \mathbb{R}$ vor Beobachtung der Daten bei gegebenem τ zu. Daher hängt die gemeinsame a-priori Verteilung von (μ, τ) nur noch von der a-priori Verteilung von τ ab. Diese sei wiederum eine Gleichverteilung auf allen Werten größer 0. Diese spezielle impropere a-priori Verteilung führt zu einer properen a-posteriori Verteilung wie wir im Verlauf dieses Kapitels noch sehen werden. Für die Parameterebene (2.23) können wir die Produktverteilungen wie in Gleichung (2.18) und (2.19) angeben

$$p(\theta_1, \dots, \theta_H|\mu, \tau) = \prod_{h=1}^H N_{\theta_h}(\mu, \tau^2)$$

¹ Die einzelnen Formeln und Zwischenschritte können im Detail in [22, Kap. 5.4] nachgeschlagen werden. Wir verzichten daher bei jeder Formel erneut auf diese Quelle zu verweisen.

$$p(\theta_1, \dots, \theta_H) = \int \prod_{h=1}^H N_{\theta_h}(\mu, \tau^2) p(\mu, \tau) d(\mu, \tau). \quad (2.25)$$

Der Mittelwert jeder Gruppe sei $\bar{y}_{.h} := \frac{1}{n_h} \sum_{i=1}^{n_h} y_{ih}$ mit Stichprobenvarianz $\sigma_h^2 = \frac{\sigma}{n_h}$. Dieser Mittelwert entspricht in unseren späteren Anwendungen der standardisierten Mittelwertdifferenz oder dem beobachteten Log Odds Ratio in der klinischen Studie h . Damit können wir die Likelihood umschreiben zu

$$y_{.h} | \theta_h \sim N(\theta_h, \sigma_h^2). \quad (2.26)$$

Nachdem wir nun das Modell spezifiziert haben, folgen wir den allgemeinen Schritten 1.-3. (vgl. Kap. 2.4) zur Bestimmung der gemeinsamen a-posteriori Verteilung aller Gruppen- und Hyperparameter.

Schritt 1. Im ersten Schritt faktorisieren wir diese Verteilung und können die Gestalt der einzelnen Verteilungen in Formel (vgl. (2.20)) genauer angeben

$$\begin{aligned} p(\theta_1, \dots, \theta_H, \mu, \tau | y) &\propto p(\mu, \tau) p(\theta_1, \dots, \theta_H | \mu, \tau) p(y | \theta_1, \dots, \theta_H) \\ &\propto p(\mu, \tau) \prod_{h=1}^H N(\theta_h | \mu, \tau^2) \prod_{h=1}^H N(\bar{y}_{.h} | \theta_h, \sigma_h^2). \end{aligned} \quad (2.27)$$

Schritt 2. Als Nächstes bestimmen wir die bedingte a-posteriori Verteilung der Gruppenparameter $p(\theta | \mu, \tau, y)$. Die Parameter θ_h sind einerseits bedingt auf μ, τ unabhängig in der a-priori Verteilung (vgl. (2.23)) und treten andererseits als Faktoren in der Likelihood auf (vgl. 2.22). Damit lässt sich die bedingte a-posteriori Verteilung $p(\theta | \mu, \tau, y)$ in H Faktoren zerlegen. Bedingt auf die Hyperparameter befinden wir uns in der Situation von Kapitel 2.2.1. Wir haben H unabhängige unbekannte a-priori normalverteilte Mittelwerte, sodass wir für jeden einzelnen Mittelwert die a-posteriori Verteilung wie in Kapitel 2.2.1 bestimmen können. Da die θ_h gegeben μ, τ unabhängig sind, gilt nach Formel (2.6)

$$\theta_h | \mu, \tau, y \sim N(\hat{\mu}, V_h) \quad (2.28)$$

$$\text{mit} \quad \hat{\mu} = \frac{\frac{1}{\tau^2} \mu + \frac{1}{\sigma_h^2} \bar{y}_{.h}}{\frac{1}{\tau^2} + \frac{1}{\sigma_h^2}} \quad \text{und} \quad \frac{1}{V_h} = \frac{1}{\tau^2} + \frac{1}{\sigma_h^2}. \quad (2.29)$$

Der a-posteriori Erwartungswert ist wie bereits zuvor ein durch die Präzisionen gewichtetes Mittel zwischen dem Stichprobenmittelwert $\bar{y}_{.h}$ in Gruppe h und dem a-priori Erwartungswert μ .

Schritt 3. Der letzte Schritt ist von der Berechnung ein wenig aufwendiger, sodass wir nur die Resultate angeben. Das Ziel besteht darin für jeden Hyperparameter die a-posteriori Verteilung zu bestimmen. In diesem Beispiel können wir sogar direkt die Gestalt der gemeinsame a-posteriori Randverteilung der Hyperparameter $p(\mu, \tau|y)$ bestimmen. Im Anschluss können wir sie für die einzelnen Hyperparameter zerlegen und erhalten so die a-posteriori Schätzungen für μ und τ . Die a-posteriori Randverteilung der Hyperparameter $p(\mu, \tau|y)$ könnten wir beispielsweise durch Integration oder analytische Berechnung aus der gemeinsamen a-posteriori Verteilung $p(\theta, \mu, \tau)$ bestimmen. Im hierarchischen Normal-Normal-Modell können wir die Informationen über die Hyperparameter aber direkt aus den Daten eruieren und die Parameterebene überspringen. Denn in der Zerlegung

$$p(\mu, \tau|y) \propto p(\mu, \tau)p(y|\mu, \tau) \quad (2.30)$$

in die gemeinsame a-priori Verteilung der Hyperparameter, die wir bereits kennen (vgl. (2.24)), und die Randverteilung der Likelihood, können wir die Verteilung $p(y|\mu, \tau)$ konkret bestimmen. Im Allgemeinen besitzt diese Verteilung keine geschlossene Form. Hier ist sie aber von vergleichsweise einfacher Form. Die Randverteilungen der Gruppenmittelwerte $\bar{y}_{.h}$ gemittelt über alle θ sind unabhängige jedoch nicht identische Normalverteilungen

$$\bar{y}_{.h}|\mu, \tau \sim N(\mu, \sigma_h^2 + \tau^2). \quad (2.31)$$

Damit können wir die marginale Verteilung schreiben als

$$p(\mu, \tau|y) \propto p(\mu, \tau)p(y|\mu, \tau) = p(\mu, \tau) \prod_{h=1}^H p(\bar{y}_{.h}|\mu, \tau) = p(\mu, \tau) \prod_{h=1}^H N_{\bar{y}_{.h}}(\mu, \sigma_h^2 + \tau^2). \quad (2.32)$$

Da nun die a-posteriori Randverteilung der Hyperparameter bekannt ist, können wir jetzt auch die a-posteriori Verteilungen für jeden Hyperparameter einzeln bestimmen.

Zunächst berechnen wir die a-posteriori Verteilung von μ bei gegebenen τ . Dafür zerlegen wir die gemeinsame Verteilung weiter in

$$p(\mu, \tau|y) = p(\mu|\tau, y)p(\tau|y). \quad (2.33)$$

Dabei ist $p(\mu|\tau, y)$ die a-posteriori Verteilung von μ bei gegebenem τ . Durch Umformung, der Verwendung einer a-priori Gleichverteilung für $p(\mu|\tau)$ im Sinne von Gleichung (2.24), der

Bedingung, dass $p(\mu, \tau|y)$ die Form in Gleichung (2.32) haben muss, sowie der Interpretation von $\bar{y}_{.h}$ als H unabhängige Schätzungen von μ erhalten wir

$$\mu|\tau, y \sim N(\hat{\mu}, V_\mu) \quad (2.34)$$

$$\text{mit } \hat{\mu} = \frac{\sum_{h=1}^H w_h \bar{y}_{.h}}{\sum_{h=1}^H w_h}, \quad V_\mu = \frac{1}{\sum_{h=1}^H w_h} \quad (2.35)$$

$$\text{und } w_h = \frac{1}{\sigma_h^2 + \tau^2}. \quad (2.36)$$

Der a-posteriori Erwartungswert $\mathbb{E}(\mu|\tau, y)$ ist damit der durch die inverse Varianz gewichtete Mittelwert $\bar{y}_{.h}$ aller Beobachtungen $y_{.h}$. Die Gewichte w_h entsprechen den aus der klassischen Metaanalyse bekannten inversen Varianzgewichten [50, S. 8; 51, S. 335]. Diese setzen sich aus der Varianz σ_h^2 innerhalb der entsprechenden Gruppe und der Varianz τ^2 zwischen allen Gruppen von Beobachtungen zusammen. Bei uns sind diese Gruppen später die einzelnen Studien, sodass dort die a-posteriori Varianz von der Interstudienvarianz τ^2 beeinflusst wird. Da die a-posteriori Verteilung $p(\mu|\tau, y)$ eine Normalverteilung ist, ist sie trotz der Anwendung von impropere a-priori Verteilungen proper.

Die letzte unbekannte Verteilung ist die a-posteriori Verteilung von τ . Indem wir die Gleichung (2.33) nach $p(\tau|y)$ auflösen und für die Verteilungen die Ergebnisse aus Gleichung (2.32) und (2.34) einsetzen, erhalten wir

$$p(\tau|y) = \frac{p(\mu, \tau|y)}{p(\mu|\tau, y)} \propto \frac{p(\tau) \prod_{h=1}^H N_{\bar{y}_{.h}}(\mu, \sigma_h^2 + \tau^2)}{N_\mu(\hat{\mu}, V_\mu)}.$$

Diese Beziehung muss für jeden Wert für μ gelten. Insbesondere müssen sich dadurch alle möglichen Vorfaktoren von μ gegenseitig wegekürzen. Der Ausdruck muss demnach auch gelten, wenn wir μ durch $\hat{\mu}$ ersetzen. Damit gilt

$$\begin{aligned} p(\tau|y) &\propto \frac{p(\tau) \prod_{h=1}^H N_{\bar{y}_{.h}}(\hat{\mu}, \sigma_h^2 + \tau^2)}{N_{\hat{\mu}}(\hat{\mu}, V_\mu)} \\ &\propto p(\tau) \sqrt{V_\mu} \prod_{h=1}^H \frac{1}{\sqrt{(\sigma_h^2 + \tau^2)}} \exp\left(-\frac{(\bar{y}_{.h} - \hat{\mu})^2}{2(\sigma_h^2 + \tau^2)}\right), \end{aligned} \quad (2.37)$$

mit $\hat{\mu}$ und V_μ aus Gleichung (2.35). Selbst in diesem einfachen Normal-Normal-Modell ist die Verteilung $p(\tau|y)$ bereits eine komplizierte Funktion in τ . In komplexeren Modellen ist die Bestimmung dieser Verteilungen schwieriger, sodass häufig MCMC-Verfahren angewandt werden.

Berechnung durch Simulationen. Für das hierarchische Normal-Normal-Modell können wir die a-posteriori Verteilung von θ durch Simulationen berechnen. Dafür benutzen wir die folgende Faktorisierung der gemeinsamen a-posteriori Verteilung der Hyperparameter und der Gruppenparameter

$$p(\theta, \mu, \tau) \propto p(\theta|\mu, \tau, y)p(\mu|\tau, y)p(\tau|y). \quad (2.38)$$

Die Verteilungen auf der rechten Seite haben die Form (2.28), (2.34) bzw. (2.37). Der Simulationsablauf ist ähnlich zum Gibbs-Sampling 2.21. Für jedes $t \in \{1, \dots, T\}$ wiederholen wir die Schritte 1-3:

1. Simuliere $\tau^{(t)} \sim p(\tau|y)$
2. Simuliere $\mu^{(t)} \sim p(\mu|\tau^{(t)}, y)$
3. Simuliere $\theta_h^{(t)} \sim p(\theta_h|\mu^{(t)}, \tau^{(t)}, y)$ mit $h = 1, \dots, H$. (2.39)

Den ersten Schritt können wir numerisch mit der inversen kumulativen Verteilungsfunktions-Methode [22, S. 25] bestimmen. Die übrigen Simulationsschritte sind dann nur noch Ziehungen aus den zugehörigen Normalverteilungen, in die jeweils die generierten Werte aus dem vorangegangenen Schritt eingesetzt werden.

Für unsere späteren Anwendungen interessiert uns speziell die a-posteriori prädiktive Verteilung von θ^* . Bei bekanntem τ ist die a-posteriori Verteilung $p(\mu|\tau, y)$ die Normalverteilung $N(\hat{\mu}, V_\mu)$ mit den Parametern in (2.35). Die prädiktive Verteilung lässt sich mit den Ergebnissen aus dem Normal-Modell (Kap. 2.2.1) und der dort dargestellten a-posteriori prädiktiven Verteilung (Formel (2.9)) herleiten. Nach [50, S. 8] besitzt sie daher bei gegebenem τ die Form

$$\theta^*|y, \tau = \theta^*|\bar{y}_{.1}, \dots, \bar{y}_{.H}, \tau \sim N\left(\frac{\sum_{h=1}^H w_h \bar{y}_{.h}}{\sum_{h=1}^H w_h}, \frac{1}{\sum_{h=1}^H w_h} + \tau^2\right)$$

mit $w_h = \frac{1}{\sigma_h^2 + \tau^2}$. (2.40)

Die a-posteriori prädiktive Verteilung ist also ebenfalls eine Normalverteilung. Sie unterscheidet sich von der a-posteriori Verteilung (2.34) nur durch die erhöhte Varianz. Die Varianz besteht aus der Summe der a-posteriori Varianz von μ und der Varianz zwischen den einzelnen Gruppen bzw. Studien τ^2 . Dies ist für die spätere meta-analytisch prädiktive Anwendung sehr zweckmäßig. Denn selbst wenn wir den mittleren wahren Behandlungseffekt den Hyperparameter μ kennen, aber zulassen, dass τ groß sein kann, also die Studien sehr heterogen

sind, kann der zukünftige wahre Studieneffekt θ^* nur sehr unpräzise vorhergesagt werden (siehe Beispiel 4.2).

2.5 Auswahl der a-priori Verteilungen

2.5.1. Kritik an der Auswahl

Personen sind in ihrer Meinung immer subjektiv beeinflusst. Insbesondere gilt dies bei der Planung klinischer Studien. *Altman* [2] kritisiert, dass das a-priori Wissen zu stark gewichtet wird, anstatt dass ein geeignetes Design gewählt wird. Zudem neigen unterschiedliche Personengruppen wie Kliniker, Vertreter von Pharmazieunternehmen oder Statistiker nach Begutachtung der Literatur zu verschiedenen Schlussfolgerungen über die Annahme des Studieneffektes der neuen Studie. Kliniker und Pharmavertreter tendieren eher zu einer Überschätzung des Behandlungseffektes, während Statistiker diesem häufig skeptischer gegenüberstehen [62, S. 147]. Desweiteren kritisieren *Bland et al.* [9, S. 317], dass die Wahl der a-priori Verteilung subjektiv ist und zu verschiedenen Entscheidungen führen kann:

„A major difficulty, of course, is deciding on the prior distribution. This is going to influence the conclusions of the study, yet it may be a subjective synthesis of the available information, so the same data analysed by different investigators could lead to different conclusions.“

Zeitliche Differenzen zwischen den vergangenen und der neuen Studie können ebenfalls eine Verfälschung der informativen a-priori Verteilung, die aus den historischen Studien gewonnen wird, bewirken. Die folgende Herangehensweise geht auf *Laplace* zurück:

„When nothing is known about θ in advance, let the prior $\pi(\theta)$ be a uniform distribution, that is, let all possible outcomes of θ have the same probability. This is also known as the principle of insufficient reason.“ [64, S. 1].

Aufgrund der Vielzahl von Kritikpunkten an der Auswahl einer geeigneten a-priori Verteilung für die Zielgröße in einer neuen klinischen Studie, ist es von enormer Wichtigkeit, ein soweit wie möglich objektives Kriterium zur Beurteilung der Evidenz historischer Daten zu wählen. Wenn die a-priori Verteilung der zukünftigen Studie auf der Basis historischer Ergebnisse gewählt wird, müssen wir darauf achten, die Unsicherheit der früheren Ergebnisse

mit zu modellieren. Der in Kapitel 3 beschriebene meta-analytisch prädiktive Ansatz stellt ein weitgehend objektives Bewertungswerkzeug der historischen Studien dar.

Besonders die Wahl der hyper a-priori Verteilungen für die Varianzparameter wurde bereits in mehreren Arbeiten untersucht und teils kritisiert [2, 11, 21, 33, 40]. Damit die Beobachtungen großen Einfluss auf die a-posteriori Verteilungen haben, werden oftmals nicht-informative a-priori Verteilungen benutzt. Die amerikanische Gesundheitsbehörde *FDA* rät explizit zur Anwendung nicht-informativer a-priori Verteilungen für den Varianzparameter zwischen den Studien:

„[...] in a Bayesian hierarchical model for combining studies, a non-informative prior distribution may be placed on a parameter that captures the variability between studies because, ordinarily, no informative prior is available on this parameter.“ [67, S. 25].

Gleichzeitig weist sie auf das Problem hin, dass die wahre Interstudienvarianz zwischen realen klinischen Studien üblicherweise unbekannt ist. Wir können uns aber mit konstruierten Simulationsszenarien behelfen, in denen wir die erwartete Variation vorgeben und somit die Performance verschiedener a-priori Verteilungen beurteilen können (vgl. Kap. 5). Während die *FDA* nicht-informative Verteilungen empfiehlt, konstatieren *Irony et al.* [34], dass es so etwas wie nicht-informative Prior nicht gibt. Bei der Verwendung von nicht-informativen Verteilungen können nämlich in speziellen Situationen Probleme auftreten:

„In a hierarchical model $\theta \sim N[\mu, \tau^2]$, the random-effects standard deviation τ plays an important role, and its value can be very influential in assessing the uncertainty concerning μ or in predicting future θ s. However, there may be limited information in the data to provide a precise estimate of τ due either to there being few units, or to each unit providing little information, or both. This can make the prior for τ particularly important, and yet neither is there any generally accepted reference prior for τ , nor are there formally established techniques for assessing a subjective prior distribution.“ [62, S. 168]

„[...] the choice of “noninformative” prior distribution can have a big effect on inferences, especially for problems where the number of groups J is small or the group-level variance σ_α^2 is close to zero.“ [21, S. 516]

Wenn nur wenige frühere Studienergebnisse vorhanden sind, können die nicht-informativen a-priori Verteilungen schnell einen sehr informativen Einfluss auf die a-posteriori Verteilung haben und diese dominieren. Der Einfluss der Daten ist dann gering. Genau diesen Zustand prüfen wir in den Simulationen in Kapitel 5.

2.5.2. Hyper a-priori Verteilungen

Auf die Wahl der passenden a-priori Verteilungen für die Hyperparameter μ und τ der hierarchischen Modelle (vgl. 3.2) legen wir in dieser Arbeit einen besonderen Schwerpunkt. Insbesondere untersuchen wir vier verschiedene hyper a-priori Verteilungen für die Gruppenvariation, hier die Interstudienvarianz, die in der Literatur empfohlene Verteilungsfamilien repräsentieren. Als Hyperverteilungen werden häufig nicht-informative a-priori Verteilungen gewählt. Das Ziel dieser Wahl ist es, die historischen empirischen Effekte hervorzuheben. Der Einfluss der Daten in Form der Likelihood soll in der a-posteriori (2.2) bzw. a-posteriori prä-diktiven Verteilung (2.4) dadurch erhöht werden. Dies kann zu den oben erwähnten Schwierigkeiten führen, insbesondere in Situationen, in denen nur wenige historische Studien vorhanden sind. In diesem Fall können sich die empirischen Daten in Form der Likelihood nicht gegen den Einfluss der a-priori Verteilung in der a-posteriori Verteilung durchsetzen. Die a-posteriori Verteilung wird somit durch die nicht-informative a-priori Verteilung dominiert. Dies kann zu verfälschten Schätzungen der einzelnen Parameter, insbesondere der Überdeckungswahrscheinlichkeiten bzw. der Intervallschätzungen führen (vgl. Bspl. 4.3). Für den Hyperparameter μ nehmen wir in dieser Arbeit an, dass er a-priori normalverteilt $N(0, 10^6)$ ist. Dies lässt eine große Streuung des erwarteten mittleren Behandlungseffekts zu. Es sei f eine Dichtefunktion $f : \mathbb{R} \rightarrow [0, \infty)$.

Normalverteilung: Die allgemeine Dichte der Normalverteilung ist

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right).$$

Für $N(0, 10^6)$ ist sie

$$f(x; 0, 10^6) = \frac{1}{10^6\sqrt{2\pi}} \exp\left(-\frac{x^2}{2 \cdot 10^{12}}\right). \quad (2.41)$$

Wir konzentrieren uns vorrangig auf die Untersuchung des Verhaltens verschiedener a-priori Verteilungen für die Variation zwischen den Studien, die Interstudienstandardabweichung τ bzw. Interstudienvarianz τ^2 . Dieser Parameter beeinflusst maßgeblich die Präzision der Vorhersage eines zukünftigen Studieneffekts und somit die Bestimmung des Gewichtungsfaktors. *Gelman* [21] betrachtet unterschiedliche Verteilungsfamilien als a-priori Verteilung für Varianzparameter. Er beschreibt folgendes generelles Vorgehen zur Wahl der geeigneten Verteilung [vgl. 21, Kap. 7.1]: Wenn es die Daten zulassen, soll möglichst eine nicht-informative

Gleichverteilung gewählt werden. Ist dies nicht möglich, soll die Auswahl der Situation angepasst werden. Beispiele einer solchen Verteilung sind eine Gleichverteilung mit großer Streubreite wie $U(0, 100)$ oder eine Half-Normal-Verteilung mit großem Skalierungsparameter, zum Beispiel $HN(0, 100)$. Falls die a-priori Information immer noch zu schwach ist, um unrealistische Werte des Varianzparameters τ^2 zu unterdrücken, soll auf die Familie der gefalteten nichtzentralen t-Verteilungen zurückgegriffen werden. Ein Spezialfall dieser Familie ist die Half-Cauchy-Verteilung, die wir als Vertreter dieser Verteilungsfamilie untersuchen. Der Vorteil der Half-Normal-Verteilung gegenüber der Inverse-Gamma-Verteilung besteht darin, dass erstere bei Schätzungen des Varianzparameters nahe bei 0 besser geeignet ist. *Lambert et al.* [40] betrachten ebenfalls eine große Auswahl von a-priori Verteilungen für Varianzparameter. Dabei stellen sie fest, dass wenn die wahre Interstudienstandardabweichung, dort „*between unit standard deviation*“ [40, S. 2519] genannt, in der Nähe von 0 liegt, Probleme bei der Konvergenz der Markovketten auftreten können. Der MCMC-Mechanismus ist gezwungen, bei jeder Ziehung positive Werte für die Standardabweichung zu generieren. Beispielsweise führt die Anwendung der ursprünglich nicht-informativen Inverse-Gamma-Verteilung $\Gamma(\epsilon, \epsilon)$ mit $\epsilon \rightarrow 0$ in dieser Situation zu einer informativen Schätzung, die stark von der Wahl des Parameters ϵ abhängt. Dies kann zu einer improporen, a-posteriori Verteilung führen. Das Integral über die nicht-normalisierte a-posteriori Dichte ist uneigentlich, sodass damit die Eigenschaft einer Wahrscheinlichkeitsverteilung nicht mehr erfüllt ist. Insbesondere können wir dann keine zulässigen Aussagen über die a-posteriori Parameter treffen:

„If the unnormalized posterior density, obtained by multiplying the likelihood by a ‘formal’ prior density representing an improper prior distribution, integrates to infinity, then the asymptotic results, which rely on probabilities summing to 1, do not follow.“ [22, S. 109f.]

Gelman schlägt vor, dass in solchen speziellen Situationen schwach informative a-priori Verteilungen angewendet werden sollen:

„In addition, when regression predictors must be estimated, more than $J = 3$ groups may be necessary to estimate σ_a [hier τ ; Anm. d. V.] from a noninformative prior distribution, thus requiring at least weakly informative prior distributions for the regression coefficients, the variance parameters, or both.“ [21, S. 529].

Dies ist im dritten klinischen Beispiel 4.3 und den betrachteten Simulationsszenarien in Kapitel 5 der Fall.

Aufgrund der vorherigen Überlegungen untersuchen wir das Verhalten folgender hyper a-priori Verteilungen für die Interstudienvarianz τ^2 bzw. die Interstudienstandardabweichung τ :

(a) Gleichverteilung: Die Dichte der Gleichverteilung lautet

$$f(x; a, b) = \begin{cases} \frac{1}{b-a} & \text{für } a \leq x \leq b \\ 0 & \text{für } x < a \vee b < x. \end{cases}$$

Wie in [21] empfohlen, verwenden wir hier eine schwach-informative Gleichverteilung mit $a = 0$ und $b = 100$. Diese Wahl entspricht immer noch einer sehr großen zulässigen Variation für τ , falls die Zielgröße eine standardisierte Mittelwertdifferenz oder ein Log Odds Ratio ist. Für die Zielgrößen in unseren Beispielen erwarten wir vielmehr Werte von τ im Bereich $[0, 1]$. Die Dichtefunktion der $U(0, 100)$ Verteilung hat die Gestalt

$$f(x) = \begin{cases} \frac{1}{100} & \text{für } 0 \leq x \leq 100 \\ 0 & \text{für } x < 0 \vee 100 < x. \end{cases} \quad (2.42)$$

(b) Inverse-Gamma-Verteilung: Eine standardmäßige Wahl der a-priori Verteilung für Varianzparameter ist in vielen Fällen die Inverse-Gamma-Verteilung

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(-\frac{\beta}{x}\right) \quad \text{für } x > 0.$$

Die Anwendung dieser Verteilung ist historisch gewachsen. Sie gehört zur Klasse der konjugierten Verteilungen. Konjugiert bedeutet, dass die a-posteriori Verteilung wieder zu derselben Verteilungsfamilie wie die a-priori Verteilung gehört, also in diesem Fall wieder eine Inverse-Gamma-Verteilung ist. Der Vorteil dieser Verteilungsfamilien liegt darin, dass die a-posteriori Verteilung in gewissen Fällen analytisch bestimmt werden kann. Wir untersuchen hier das Verhalten der Inverse-Gamma-Verteilung mit den Parametern $\alpha = 0.001$ und $\beta = 0.001$ für die Interstudienvarianz τ^2 , also $\tau^2 \sim \Gamma^{-1}(0.001, 0.001)$,

$$f(x) = \frac{0.993116}{\Gamma(0.001)} x^{-1.001} \exp\left(-\frac{0.001}{x}\right) \quad \text{für } x > 0. \quad (2.43)$$

$\Gamma^{-1}(0.001, 0.001)$ ist eine der gebräuchlichsten a-priori Verteilungen, was wahrscheinlich daran liegt, dass sie in vielen Beispielen der Softwarepakete WinBUGS [43] und

OpenBUGS [44] Anwendung findet [60, 61]. In der Regel gilt diese spezielle Inverse-Gamma-Verteilung als nicht-informativ. *Spiegelhalter et al.* [62, S. 170] bezeichnen diese Verteilung als „just proper“. Sie ist approximativ eine Gleichverteilung für $\log(\tau)$ über nahezu den gesamten Wertebereich. Allerdings besitzt sie in der Nähe von 0 einen, wenn auch kleinen, Spike. Das heißt, in einem kleinen Bereich der Verteilung befindet sich relativ viel Masse. Falls die Likelihood kleine Werte für τ liefert, so unterstützt $\Gamma^{-1}(0.001, 0.001)$ vorrangig eine kleine Varianz. Wenn also bei den MCMC-Simulationen Werte in der Nähe von 0 generiert werden und die Likelihood ebenfalls eine kleine Varianz begünstigt, kann aus der nicht-informativen Verteilung eine informative a-priori Verteilung werden [vgl. 21, 40]. Da die Iterationen der MCMC-Simulationen nicht unabhängig gezogen, sondern jeweils bedingt auf die letzte Iteration generiert werden, kann dies dazu führen, dass die Markov-Kette im Bereich des Spikes stecken bleibt. Möglicherweise wird dann die Varianz unterschätzt und die Markov-Kette konvergiert im schlimmsten Fall nicht gegen die wahre stationäre Verteilung.

Für die spätere Programmierung der Verteilung in OpenBUGS benutzen wir die folgende Beziehung zwischen der Inverse-Gamma und der Gamma-Verteilung:

Satz 2.1. *Sei X eine absolut stetige Zufallsvariablen mit $X \sim \Gamma(\alpha, \beta)$, dann ist die Zufallsvariable $Y := \frac{1}{X} \sim \Gamma^{-1}\left(\alpha, \frac{1}{\beta}\right)$ verteilt.*

Beweis. X besitzt die Dichte

$$f_X(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right).$$

Da die Γ - und Γ^{-1} -Verteilung nur für Werte größer Null definiert ist, ist die Transformation $1/x$ für $x > 0$ eine bijektive stetig differenzierbare Abbildung mit stetig differenzierbarer Umkehrfunktion. Mit dem Transformationssatz für Dichten [15, S. 191] erhalten wir

$$\begin{aligned} f_Y(y) &= f_X\left(\frac{1}{y}\right) \cdot \left|\frac{d}{dy}(y^{-1})\right| \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha} y^{-(\alpha-1)} \exp\left(-\frac{y^{-1}}{\beta}\right) \cdot y^{-2} \\ &= \frac{\left(\frac{1}{\beta}\right)^\alpha}{\Gamma(\alpha)} y^{-\alpha-1} \exp\left(-\frac{1}{\beta y}\right). \end{aligned}$$

□

Bemerkung 2.1. An dieser Stelle weisen wir darauf hin, dass die Inverse-Gamma-Verteilung in den hierarchischen Modellen dieser Arbeit als hyper a-priori Verteilung für die Varianz τ^2 und nicht für die Standardabweichung τ verwendet wird. Alle anderen Startverteilungen beziehen sich auf τ . Wenn wir im weiteren Verlauf von der Inverse-Gamma-Verteilung als a-priori Verteilung für τ sprechen, ist damit immer gemeint, dass sich die Startverteilung auf die Varianz τ^2 bezieht.

(c) **Half-Normal-Verteilung:** Diese Verteilung ist ein Spezialfall der gefalteten Normalverteilung. Die Dichtefunktion der gefalteten Normalverteilung ist

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) + \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(-x-\mu)^2}{2\sigma^2}\right) \quad \text{für } x \geq 0.$$

Die Half-Normal-Verteilung ergibt sich, wenn die Faltung genau bei dem x liegt, für welches die kumulative Wahrscheinlichkeit $p = 0.5$ beträgt. Der Parameter μ bezeichnet hier nicht den Erwartungswert, sondern den Faltungspunkt der Verteilung. Mit $\mu = 0$ und Skalierungsparameter $\sigma^2 = 1$ ergibt sich somit die Dichte der Half-Normal-Verteilung $HN(0, 1)$

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \quad \text{für } x > 0. \quad (2.44)$$

Die $HN(0, 1)$ -Verteilung ist somit eine Standardnormalverteilung, die bei 0 zusammengefasst ist. Mit anderen Worten, falls eine Zufallsvariable X standardnormalverteilt ist, also $X \sim N(0, 1)$, so ist $|X| \sim HN(0, 1)$ verteilt. Für unsere Simulationen wählen wir später die $HN(0, 1)$ als a-priori Verteilung für die Interstudienstandardabweichung τ . Die $HN(0, 1)$ -Verteilung lässt Werte für τ größer als 2 mit einer kleinen Wahrscheinlichkeit von ca. 5% zu. Für die hier betrachteten Zielgrößen wie die Log Odds Ratios und die standardisierten Mittelwertdifferenzen entspricht dies einer sehr großen Heterogenität und breiten Streuung zwischen den einzelnen Studien. Die historischen Daten würden in diesem Fall keine bedeutende Information über den Parameter θ^* in der neuen Studie enthalten.

(d) **Half-Cauchy-Verteilung:** Die Half-Cauchy-Verteilung

$$f(x; \alpha, \beta) = \frac{2}{\pi\beta \left(1 + \left(\frac{x-\alpha}{\beta}\right)^2\right)} \quad \text{für } x \geq t$$

gehört zu den gefalteten nichtzentralen t -Verteilungen. Wir betrachten die spezielle Startverteilung $\tau \sim HC(0, 1)$ mit Dichte

$$f(x) = \frac{2}{\pi(1+x^2)} \quad \text{für } x > 0. \quad (2.45)$$

Da die Half-Cauchy-Verteilung nicht in den Programmen WinBUGS und OpenBUGS implementiert ist, werden wir auf einen Hilfsatz zurückgreifen, der es uns ermöglicht über andere Verteilungen eine Half-Cauchy-Verteilung zu konstruieren. Wir folgen dem Vorgehen in [46, S. 4ff.] und verwenden den nachstehenden Satz:

Satz 2.2. *Seien X_1 und X_2 zwei unabhängige, absolut stetige Zufallsvariablen mit $X_1 \sim N(0, \sigma_1^2)$ und $X_2 \sim N(0, \sigma_2^2)$, dann gilt*

$$\frac{X_1}{X_2} \sim \text{Cauchy}\left(\frac{\sigma_1}{\sigma_2}\right).$$

Beweis. Für alle $z, t \in \mathbb{R}$ gilt mit der Formel für die Dichtefunktion das Produkt unabhängiger Zufallsvariablen [14, S. 55]

$$\begin{aligned} f_{X_1/X_2}(z) &= \int_{-\infty}^{\infty} |t| f_{X_1}(z \cdot t) f_{X_2}(t) dt \\ &= \int_{-\infty}^{\infty} |t| \frac{1}{\sqrt{2\pi\sigma_1^2}} \cdot \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{t^2 z^2}{2\sigma_1^2} - \frac{t^2}{2\sigma_2^2}\right) dt \\ &\stackrel{\text{Sym.}}{=} \int_0^{\infty} 2t \frac{1}{\sqrt{2\pi\sigma_1^2}} \cdot \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{t^2}{2} \underbrace{\left(\frac{z^2}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)}_{=: \nu}\right) dt \\ &\text{Substitution mit } dt = \frac{1}{t} \left(\frac{\sigma_1^2 \sigma_2^2}{\sigma_2^2 z^2 + \sigma_1^2}\right) d\nu \\ &\stackrel{\text{Subst.}}{=} \int_0^{\infty} \frac{1}{\pi \sigma_1 \sigma_2} \cdot \frac{\sigma_1^2 \sigma_2^2}{\sigma_2^2 z^2 + \sigma_1^2} \exp(-\nu) d\nu \\ &= \frac{1}{\pi} \cdot \frac{\sigma_1 \sigma_2}{\sigma_2^2 z^2 + \sigma_1^2} \underbrace{\int_0^{\infty} \exp(-\nu) d\nu}_{=1} \\ &= \frac{1}{\pi} \cdot \frac{\sigma_1 \sigma_2}{\sigma_2^2 z^2 + \sigma_1^2} \\ &= \frac{1}{\pi} \cdot \frac{\frac{\sigma_1}{\sigma_2}}{z^2 + \left(\frac{\sigma_1}{\sigma_2}\right)^2} \\ &\stackrel{b:=\frac{\sigma_1}{\sigma_2}}{=} \frac{1}{\pi} \cdot \frac{b}{z^2 + b^2}. \end{aligned} \quad (2.46)$$

□

Insbesondere folgt aus Satz 2.2, wenn wir zwei unabhängige standardnormalverteilte Zufallsvariablen $X_1 \sim N(0,1)$ und $X_2 \sim N(0,1)$ generieren, dass

$$\left| \frac{X_1}{X_2} \right| \sim HC(0,1) \quad (2.47)$$

gilt. Die Programmierung der Half-Cauchy-Verteilung in WinBUGS bzw. OpenBUGS setzen wir analog zu *Marley et al.* [46, S. 6] um.

3 Meta-analytisch prädiktiver Ansatz

Wir kommen nun zum Kern dieser Arbeit, einem meta-analytisch prädiktiven Ansatz zur objektiven Bestimmung des Informationsgehalts von historischen Studien für den Behandlungseffekt einer zukünftigen Studie. Im letzten Abschnitt haben wir einige grundlegende Definitionen und Beispiele der Bayes-Statistik besprochen und eine Auswahl an a-priori Verteilungen kennengelernt. Damit haben wir die Rahmenbedingungen geschaffen, um die hierarchischen Modelle innerhalb dieses Ansatzes zu verstehen und genauer zu analysieren.

3.1 Historischer Überblick der Methodik

Die Extrapolation von historischen Studienergebnissen mit Hilfe von Bayes-Verfahren wurde von mehreren Autoren untersucht und ständig weiterentwickelt. *Goodman et al.* [25] stellen im Zusammenhang mit klinischen Studien eine einfache Methode dar, wie Vorwissen aus historischen Studien als a-priori Information in die Auswertung einer neuen Studie eingebracht werden kann. Der Effekt der früheren Erwachsenenstudien wird durch eine Metaanalyse ermittelt. Für die Extrapolation auf eine neue Kinderstudie wird der gepoolte Effekt der Erwachsenen rezentriert und die inverse Varianz bzw. die Präzision um einen Faktor verkleinert. Dies soll die Unsicherheit zwischen den historischen Erwachsenenendaten und den neuen Beobachtungen an Kindern ausdrücken. Die Bestimmung des Gewichtungsfaktors erfolgt allerdings nicht empirisch, sondern hängt von der subjektiven Wahl der Studienplaner ab. Eine weitere Einschränkung dieses Ansatzes besteht darin, dass er es nicht zulässt, externe Daten aus unterschiedlichen Populationen zu benutzen. Somit ist es nicht möglich, frühere Ergebnisse aus Kinder- und Erwachsenenstudien simultan zu verwenden. *Schoenfeld et al.* [58] schlagen ein Design für eine neue pädiatrische Studie vor, für das Sie historische Ergebnisse von austauschbaren Studien an Erwachsenen extrapolieren. Sie benutzen ein hierarchisches Bayes-Modell. Darin berücksichtigen sie die Unsicherheit für die Übertragung der Daten von Erwachsenen auf Kinder. Die Extrapolation beinhaltet einen Gewichtungsfaktor, der wiederum in eine gewisse Anzahl von Patienten umgerechnet werden kann. Die Informationen der Erwachsenen in Form der Präzision des Behandlungseffekts werden herabgewichtet und verlieren somit an Wert. Die gewichteten, gepoolten Ergebnisse fließen anschließend in die Auswertung der

Kinderstudie ein. Dieser Ansatz ist allerdings ebenso wie der vorherige Versuch [25] eingeschränkt, da der Gewichtungsfaktor subjektiv, beispielsweise auf Basis von Expertenwissen, festgelegt wird.

Eine empirische Bestimmung des Gewichtungsfaktors erfolgt erstmals im Ansatz von *Neuenschwander et al.* [50]. Dort wird nicht nur die a-posteriori Verteilung des gepoolten Effekts von historischen Kontrollen geschätzt, sondern über einen prädiktiven Ansatz auch die a-posteriori prädiktive Verteilung eines zukünftigen Behandlungseffekts. Dadurch erfolgt die Extrapolation der Information weitgehend objektiv. Der Gewichtungsfaktor wird empirisch ermittelt (3.6). Er errechnet sich aus dem Verhältnis der Präzision des gepoolten Studieneffektes im vollen Modell (Kap. 3.2), in dem die Varianz zwischen den Studien als zusätzlicher Parameter geschätzt wird, und der Präzision aus dem eingeschränkten Modell (Kap. 3.3.1), in dem angenommen wird, dass die Interstudienvarianz gleich 0 ist. Der Gewichtungsfaktor wird in eine Anzahl virtueller Patienten übersetzt, die in *Neuenschwander et al.* [50] als Prior Effective Sample Size N_{eff} bezeichnet wird. Sie soll intuitiv die Anzahl der Patienten repräsentieren, die in einer zukünftigen Studie eingespart werden können, wenn die a-posteriori prädiktive Verteilung des studienspezifischen Effekts als a-priori Verteilung in der neuen Studie genutzt wird. Wir werden später zeigen, dass die so definierte Prior Effective Sample Size in *Neuenschwander et al.* [50] nicht die reelle Fallzahl ist, um die eine neue Studie reduziert werden kann. Allerdings kommt sie in speziellen Fällen sehr Nahe an die echte Anzahl heran. Die Bezeichnung Prior Effective Sample Size ist somit irreführend, da sie nur ein intuitives Maß für den Gewichtungsfaktor darstellt und nicht der tatsächlichen a-priori Information einer neuen Studie entspricht. Um Verwechslungen zu vermeiden, bezeichnen wir darum die Fallzahl in *Neuenschwander et al.* [50] mit *Precision Weight Effective Sample Size*. Ferner beschränken sich *Neuenschwander et al.* [50] auf die Extrapolation der Daten der Kontrollen. Die grundsätzliche Methodik wird in *Gerß et al.* [23, 24] aufgegriffen und erweitert. Hier wird dargestellt, wie über einen Verschiebungsparameter Informationen aus unterschiedlichen Patientenkollektiven, wie Kindern und Erwachsenen, kombiniert werden können. Der Verschiebungsparameter ist die mittlere Differenz zwischen den mittleren Behandlungseffekten der jeweiligen Studienkollektive. Zudem beschränkt sich die Extrapolation nicht länger nur auf die Daten der Kontrollgruppe, sondern wird zusätzlich auf die Ergebnisse der Behandlungsgruppe erweitert. Die mögliche Problematik bei Vorliegen weniger historischer Studien wird zwar kurz angesprochen, aber nicht genauer dargestellt und untersucht. Zudem wird in *Gerß et al.* [24] nur eine hyper a-priori Verteilung für den Varianzparameter zwischen den Studien angewandt.

In dieser Arbeit analysieren wir die Problematik bei Vorliegen weniger Vorstudien näher. Hierfür werden wir exemplarisch zeigen, in welchen Fällen der meta-analytisch prädiktive Ansatz sensitiv auf die gewählte hyper a-priori Verteilung reagiert. In diesem Rahmen betrachten wir die in Abschnitt 2.5 beschriebenen Verteilungen und beurteilen deren Anwendungsperformance. Zusätzlich geben wir eine Empfehlung für die Mindestanzahl benötigter Vorstudien, die zur stabilen Anwendung des meta-analytisch prädiktiven Ansatzes benötigt werden. Darüber hinaus werden wir zeigen, welche Auswirkungen die Parameterschätzungen auf die Power in einer neuen Studie haben. Außerdem untersuchen wir, inwiefern die Precision Weight Effective Sample Size und die Prior Effective Sample Size miteinander korrelieren.

3.2 Volles Modell

Der meta-analytisch prädiktive Ansatz nach *Neuenschwander et al.* [50] und *Gerß et al.* [24] basiert auf einem dreistufigen hierarchischen Modell. Dieses Modell bezeichnen wir als volles Modell im Gegensatz zum eingeschränkten Referenzmodell in Kapitel 3.3.1. Wir verwenden die Parameterschätzungen aus dem vollen Modell als informative a-priori Parameter in der neuen Studie. Die Gesamtfallzahl der historischen Erwachsenen- und Kinderstudien sei definiert durch

$$N_{hist} := \sum_{h=1}^{A+K} n_h.$$

Das hierarchische Modell stellt sich wie folgt dar: Wir nehmen an, die beobachteten historischen Studienresultate seien normalverteilt

$$\begin{aligned} y_a | \theta_a &\stackrel{u.i.v.}{\sim} N(\theta_a, \sigma_a^2) \\ y_k | \theta_k &\stackrel{u.i.v.}{\sim} N(\theta_k, \sigma_k^2). \end{aligned} \tag{3.1}$$

Dabei sei y_a der beobachtete Studieneffekt in der historischen Erwachsenenstudie $a \in \{1, \dots, A\}$ und y_k der beobachtete Studieneffekt in der historischen Kinderstudie $k \in \{A+1, \dots, A+K\}$ ¹. Wenn der Studieneffekt die Bedingung der Normalverteilungsannahme erfüllt, könnte dieser zum Beispiel ein Log Odds Ratio oder eine Mittelwertdifferenz sein. Die wahren Intrastudienvarianzen der Erwachsenen σ_a^2 und der Kinder σ_k^2 werden als bekannt angenommen. Die

¹ Falls keine historischen Erwachsenenstudien vorliegen sei $A = 0$.

Parameter, die die Gruppeneffekte beschreiben, also die studienspezifischen wahren Therapieeffekte, seien $\theta_a = \mathbb{E}(y_a|\theta_a)$ mit $a \in \{1, \dots, A\}$ und $\theta_k = \mathbb{E}(y_k|\theta_k)$ mit $k \in \{A+1, \dots, A+K\}$. Diese Parameter seien ebenfalls normalverteilt mit den Hyperparametern μ, μ_{shift} und τ^2

$$\begin{aligned} \theta_a|\mu &\sim N(\mu, \tau^2) && \text{mit } a \in \{1, \dots, A\} \\ \theta_k|\mu, \mu_{shift} &\sim N(\mu + \mu_{shift}, \tau^2) && \text{mit } k \in \{A+1, \dots, A+K\}. \end{aligned} \quad (3.2)$$

Hierbei ist τ^2 die Interstudienvarianz der studienspezifischen Effekte, μ der gepoolte mittlere Therapieeffekt der Erwachsenenstudien und μ_{shift} der Verschiebungsparameter zwischen dem mittleren Behandlungseffekt der Erwachsenen- und dem der Kinderpopulation. Die Parameter für die Kinder besitzen die gleiche Varianz wie die für die Erwachsenen. Sie sind nur im Mittel durch μ_{shift} verschoben.

Auf der obersten Ebene des hierarchischen Modell, der Hyperparameterebene, untersuchen wir folgende a-priori Verteilungen:

$$\begin{aligned} \mu &\sim N(0, 10^6) \\ \mu_{shift} &\sim N(0, 10^6) \end{aligned}$$

- $\tau^2 \sim \Gamma^{-1}(0.001, 0.001)$ Inverse-Gamma-Verteilung,
- $\tau \sim HN(0, 1)$ Half-Normal-Verteilung,
- $\tau \sim HC(0, 1)$ Half-Cauchy-Verteilung,
- $\tau \sim U(0, 100)$ Gleichverteilung. (3.3)

Die Dichten der hyper a-priori Verteilungen sind ausführlich in den Gleichungen (2.42), (2.43), (2.44) und (2.45) dargestellt. Im Gegensatz zum hierarchischen Normal-Normal-Modell in Kapitel 2.4.2 nehmen wir vereinfacht an, dass wir die Hyperparameter μ und τ^2 unabhängig voneinander modellieren können. Weiter betrachten wir im Gegensatz zur nicht-informativen Gleichverteilung hier eine Normalverteilung mit großer Varianz als hyper a-priori Verteilung für μ . Durch die breite Streuung der gewählten Normalverteilung besitzt sie aber einen vergleichbar niedrigen Einfluss auf die a-posteriori Verteilung wie die nicht-informative Gleichverteilung. Eine Voraussetzung für die Anwendbarkeit des vorgestellten Modells ist die Annahme, die einzelnen Studienparameter seien austauschbar im Sinne von Kapitel 2.3.1a. Das bedeutet, die studienspezifischen Effekte θ_a mit $a \in \{1, \dots, A\}$ und θ_k mit $k \in \{A+1, \dots, A+K\}$

sind Realisationen von unabhängig und $N(\mu, \tau^2)$ bzw. $N(\mu + \mu_{shift}, \tau^2)$ verteilten Zufallsvariablen (vgl. Formel (3.2)). Die Unsicherheit zwischen den beiden Kollektiven wird dadurch bedacht, dass μ_{shift} nicht fest ist, sondern ebenfalls eine Verteilung besitzt.

Die hierarchische Struktur erlaubt es über die Hyperparameter Ebene eine Verbindung zwischen den Parameterebenen der beiden Kollektive zu generieren. Dadurch kann beispielsweise der zukünftige Behandlungseffekt in jeweils einem Kollektiv auf der gemeinsamen Datenbasis beider Kollektive geschätzt werden. Die historischen Daten beider Gruppen können somit einerseits für eine neue Studie an Erwachsenen unter zusätzlicher Berücksichtigung der Kinderdaten, oder andererseits für eine neue Studie an Kindern unter zusätzlicher Berücksichtigung der Erwachsenenendaten verwendet werden. Da die Vorgehensweise in beiden Fällen analog ist, genügt es exemplarisch eines der möglichen zukünftigen Szenarien zu betrachten. Im Folgenden betrachten wir den für die klinische Praxis relevanteren Fall, die Extrapolation der früheren Ergebnisse auf eine neue pädiatrische Studie.

Das Hauptziel der bayesianischen Metaanalyse im vollen Modell ist die Bestimmung der a-posteriori prädiktiven Verteilung $p(\theta_{ped}^* | y_1, \dots, y_{A+K})$ des Studienparameters θ_{ped}^* einer zukünftigen Studie an Kindern auf der Grundlage der beobachteten Ergebnisse in den Kinder- und Erwachsenenstudien. Daraus resultiert die Varianz bzw. die Präzision von θ_{ped}^* . Die prädiktive a-posteriori Varianz von θ_{ped}^* bezeichnen wir mit

$$V_{ped, \tau, \mu_{shift}} := \text{Var}(\theta_{ped}^* | y_1, \dots, y_{A+K})^1.$$

Die Indizierung mit τ und μ_{shift} soll den Einfluss der Verteilungen des Verschiebungsparameters μ_{shift} und der Interstudienvarianz τ^2 auf die prädiktive Varianz verdeutlichen. Beide Parameter sind variabel und werden ebenfalls im Modell geschätzt. Für den mittleren Verschiebungsparameter μ_{shift} und die Interstudienvarianz τ^2 erhalten wir als Ergebnis die jeweilige a-posteriori Verteilung. Diese Parameter haben direkten Einfluss auf die a-posteriori prädiktive Verteilung von θ_{ped}^* und spielen daher auch eine zentrale Rolle bei der Berechnung des Gewichtungsfaktors und der Prior Effective Sample Size.

Die gesamte existierende Information über den zukünftigen Behandlungseffekt ist in der a-posteriori prädiktiven Verteilung von θ_{ped}^* enthalten. Um die Inflationierung der historischen Daten im vollen Modell zu quantifizieren, leiten wir analog zu *Neuenschwander et al.* [50] einen

¹ Die a-posteriori prädiktive Varianz $V_{ped, \tau, \mu_{shift}}$ hängt zusätzlich von den beobachteten Studieneffekten y_1, \dots, y_{A+K} ab. In der Bezeichnung berücksichtigen wir diese Abhängigkeit nicht, sondern weisen durch die Indizes lediglich auf den Einfluss der zusätzlich im Modell geschätzten variablen Parameter hin.

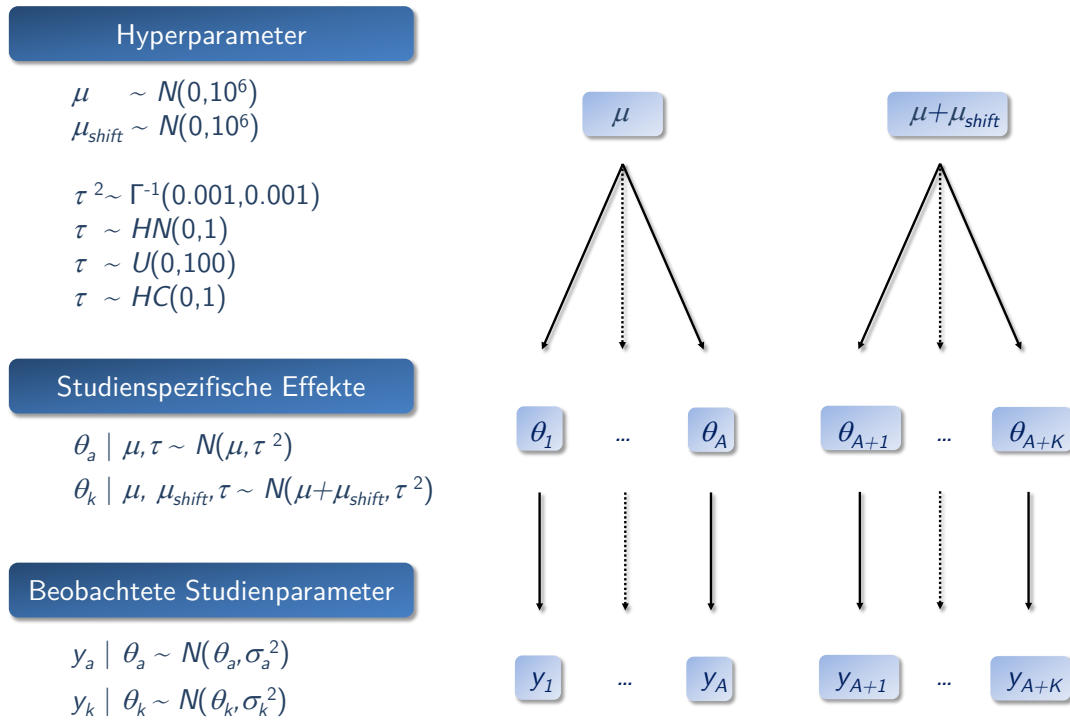


Abb. 3.1: Schema des vollen hierarchischen Modells..

Gewichtungsfaktor ω_{ped} her. Diesen werden wir zur besseren Veranschaulichung ebenfalls wie *Neuenschwander et al.* [50] in eine entsprechende Zahl virtueller Kinderpatienten $N_{ped,eff}$ umwandeln.

In diesem Modell sind die studienspezifischen Effekte per Annahme normalverteilt (vgl. (3.1), (3.2), Kap. 2.3.1b). Die entscheidende Voraussetzung zur Bestimmung von ω_{ped} ist, dass die Präzision bzw. das Inverse der Varianz proportional zum Informationsgehalt ist. Dies gilt, falls die a-posteriori prädiktive Verteilung des studienspezifischen Effekts eine Normalverteilung ist. Die Präzision gibt dann direkt die prognostische Güte der historischen Schätzungen über den zukünftigen Parameter θ_{ped}^* wieder. Im hierarchischen Normal-Normal-Modell (Kap. 2.4.2) unter den dort getroffenen Voraussetzungen und gegebener Varianz τ^2 ist die a-posteriori prädiktive Verteilung eine Normalverteilung und besitzt die Gestalt in Gleichung (2.40).

3.3 Gewichtung der Präzision

Der Wert der historischen Information ist ein Ausdruck, der nicht genau mathematisch definiert ist. Um aber das Ausmaß des Informationsverlusts der historischen Daten für die neue

Studie in einer greifbaren Form auszudrücken, leiten wir im Folgenden einen approximativen Gewichtungsfaktor her. Im vereinfachten Modell, wenn wir nur ein Kollektiv betrachten, die Interstudienvarianz τ^2 als bekannt annehmen und eine nicht-informative Gleichverteilung als hyper a-priori Verteilung für μ benutzen, befinden wir uns in der Situation des Normal-Normal-Modells in Kapitel 2.4.2. Hier ist die a-posteriori prädiktive Verteilung von θ^* (Formel (2.40)) wieder eine Normalverteilung. Bei einer Normalverteilung ist die Präzision ein Maß für die symmetrische Streuung um den Erwartungswert. Die Präzision spiegelt also in diesem Fall das Wissen über den zukünftigen Studieneffekt θ^* auf Basis der historischen Daten wieder.

In unserem hierarchischen Modell ist die Interstudienvarianz aber nicht bekannt, sondern wird als zusätzlicher Parameter mit geschätzt. Deshalb hängt die a-posteriori prädiktive Verteilung von θ_{ped}^* darüber hinaus von der Wahl der a-priori Verteilungen für τ ab. Die a-posteriori prädiktive Verteilung von θ_{ped}^* hat somit eine komplexe Gestalt und ist keine Normalverteilung. Allerdings können wir auf Grundlage einer bayesschen Interpretation des Zentralen Grenzwertsatzes [8, Theorem 2.1, S. 28] die a-posteriori prädiktive Verteilung durch eine Normalverteilung approximieren. Damit stellt die a-posteriori prädiktive Präzision $V_{ped,\tau,\mu_{shift}}^{-1}$ approximativ das historische Vorwissen über θ_{ped}^* dar.

Die prädiktive Präzision des gepoolten mittleren Therapieeffekt wird im vollen Modell unter Beachtung der Varianz zwischen den Studien τ^2 und des Verschiebungsparameters μ_{shift} für die Vorhersage des Therapieeffekt einer neuen Studie verkleinert. Interessant ist es nun zu quantifizieren, wie viel niedriger diese Präzision zur nicht inflationierten Präzision des gepoolten historischen Behandlungseffekts ist.

3.3.1. Referenzmodell

Um den Gewichtungsfaktor ω_{ped} approximativ zu bestimmen, benötigen wir ein Referenzmodell. Dieses Modell ist eine eingeschränkte Version des vollen Modells aus Kapitel 3.2. Ziel ist es, die hypothetische a-posteriori prädiktive Präzision zu bestimmen, falls die historischen Ergebnisse keinen Wertverlust erfahren würden. Dafür müssen wir uns fragen, wann genau dieser Extremfall eintritt. Die Antwort ist vergleichsweise einfach. Der Wertverlust ist genau dann minimal, wenn einerseits keine Interstudienvariation vorliegt, also $\tau^2 \equiv 0$ gilt, und andererseits der Verschiebungsparameter μ_{shift} zwischen den zwei Populationen fest und exakt bekannt ist. Wir setzen also voraus, dass die Studien in jedem Kollektiv absolut homogen sind. In dieser Situation können wir die Daten der Erwachsenen und die durch μ_{shift} ver-

schobenen Daten der Kinder gepoolt nutzen. Der gemeinsame Effekt ist dann der gewichtete Mittelwert aller Beobachtungen, wobei die Gewichtung den inversen Varianzen innerhalb der Studien entspricht. Das heißt, größere bzw. präzisere Studien bekommen mehr Gewicht als kleinere bzw. ungenauere Studien. Da die Interstudienvarianz 0 ist, stimmen im vereinfachten hierarchischen Normal-Normal-Modell die a-posteriori und die a-posteriori prädiktive Verteilung überein (vgl. Formeln (2.34),(2.40)). Der einzige Einfluss auf die Vorhersagekraft bzw. auf den Extrapolationsumfang ist somit die Intrastudienvarianz. Diese Strategie können wir als eine Fixed-Effects-Metaanalyse im klassischen Sinne interpretieren.

Unter der Annahme, dass keine Interstudienvariation vorliegt und der Verschiebungsparameter bekannt ist, berechnen wir nochmals die a-posteriori prädiktive Verteilung von θ_{ped}^* in einem zweiten hierarchischen Modell, das wir als Referenzmodell bezeichnen. Der einzige Unterschied zum vollen Modell besteht darin, dass wir für die hyper a-priori Verteilungen der Hyperparameter μ_{shift} und τ Punktmaße benutzen. Wir setzen $\tau \equiv 0$ und wählen den a-posteriori Erwartungswert von μ_{shift} aus dem vollen Modell als festen hyper a-priori Wert. Damit ergibt sich das Referenzmodell aus dem vollen Modell mit abgeänderter Hyperparameterenebene

$$\begin{aligned}\mu &\sim N(0, 10^6), \\ \mu_{shift} &= \mu_{shift_0} \equiv \mathbb{E}(\mu_{shift} | y_1, \dots, y_{A+K}), \\ \tau &= \tau_0 \equiv 0.\end{aligned}\tag{3.4}$$

Das bedeutet auf der Parameterebene, dass alle studienspezifischen Parameter mit dem jeweiligen Hyperparameter übereinstimmen

$$\begin{aligned}\theta_a | \mu &= \mu && \text{mit } a \in \{1, \dots, A\}, \\ \theta_k | \mu, \mu_{shift_0} &= \mu + \mu_{shift_0} && \text{mit } k \in \{A+1, \dots, A+K\}.\end{aligned}\tag{3.5}$$

Folglich wird die Streuung der Beobachtungen um den gemeinsamen festen Effekt nur durch die jeweiligen Intrastudienvarianzen σ_a^2 bzw. σ_k^2 beeinflusst. Mit diesem Modell erhalten wir die eingeschränkte a-posteriori prädiktive Referenzvarianz $V_{ped, \tau \equiv 0, \mu_{shift_0}}$ bei Kindern und die eingeschränkte a-posteriori prädiktive Referenzvarianz $V_{ad, \tau \equiv 0, \mu_{shift_0}}$ bei Erwachsenen. In der hypothetischen Situation ohne Inflation entspricht dieses Ergebnis dem Fall, in dem wir den gepoolten Effekt auf Grundlage beider Kollektive ohne Herabgewichtung in eine neue Studie einbringen würden. Wie im obigen realistischeren Fall, dem vollen Modell, können wir die a-

posteriori prädiktive Verteilung wieder durch eine Normalverteilung approximieren [vgl. 8, S. 28]. Damit repräsentiert die neue Präzision $V_{ped,\tau\equiv 0,\mu_{shift_0}}^{-1}$ den Wert der historischen Studien, wenn kein Informationsverlust auftritt.

Schließlich können wir den Gewichtungsfaktor angeben, der die Inflation der historischen Ergebnisse beschreibt. Er ist der Quotient der Präzision aus dem vollen Modell und der Präzision aus dem Referenzmodell bezüglich θ_{ped}^*

$$\omega_{ped} := \frac{V_{ped,\tau,\mu_{shift}}^{-1}}{V_{ped,\tau\equiv 0,\mu_{shift_0}}^{-1}}. \quad (3.6)$$

Somit drückt ω_{ped} relativ gesehen aus, wie viel Wert die gesamten früheren Studien in die neue Studie einbringen. Auf Varianzebene ausgedrückt ist der Faktor ω_{ped} einfach die a-posteriori prädiktive Varianz aus dem Referenzmodell geteilt durch die Varianz aus dem vollen Modell. Für die Extrapolation der Ergebnisse auf eine Studie in einem Erwachsenenkollektiv sei ω_{ad} der analog definierte Gewichtungsfaktor.

3.3.2. Precision Weight Effective Sample Size

Neuenschwander et al. [50] übersetzen diesen Gewichtungsfaktor in eine Anzahl virtueller Patienten. Diese bezeichnen wir als *Precision Weight Effective Sample Size*. Intuitiv soll sie die Anzahl der Patienten repräsentieren, die in einer zukünftigen Studie eingespart werden können, wenn die a-posteriori prädiktive Verteilung des studienspezifischen Effekts als a-priori Verteilung in der neuen Studie genutzt wird. Sie argumentieren damit, dass die a-posteriori prädiktive Präzision $V_{ped,\tau\equiv 0,\mu_{shift_0}}^{-1}$ im Referenzmodell (Kap. 3.3.1) die volle Information der historischen Daten widerspiegelt. Diese volle Information setzen sie der Gesamtfallzahl N_{hist} gleich. Damit ist dann $V_{ped,\tau\equiv 0,\mu_{shift_0}}^{-1}$ proportional zu N_{hist} . Die gleiche Argumentation benutzen sie für die reduzierte Information. So soll $V_{ped,\tau,\mu_{shift}}^{-1}$ proportional zur unbekannt reduzierten Fallzahl $N_{ped,eff}$ sein. Daraus schließen sie intuitiv, dass das Verhältnis der beiden Präzisionen und der beiden Fallzahlen gleich ist

$$\frac{N_{ped,eff}}{V_{ped,\tau,\mu_{shift}}^{-1}} = \frac{N_{hist}}{V_{ped,\tau\equiv 0,\mu_{shift_0}}^{-1}}.$$

Durch Umformung kann somit die Precision Weight Effective Sample Size für Kinder berechnet werden

$$N_{ped,eff} = N_{hist} \cdot \frac{V_{ped,\tau,\mu_{shift}}^{-1}}{V_{ped,\tau=0,\mu_{shift_0}}^{-1}} = N_{hist} \cdot \omega_{ped}. \quad (3.7)$$

Nach dieser Formel werden wir für alle Anwendungen in dieser Arbeit die Precision Weight Effective Sample Size berechnen. Die virtuelle Fallzahl ist also nichts anderes als die historische Gesamtfallzahl multipliziert mit dem Quotient der Präzision aus dem vollen Modell und der Präzision aus dem Referenzmodell bezüglich θ_{ped}^* . Für die Berechnung der Fallzahl $N_{ad,eff}$ für eine neue Studie an Erwachsenen müssen wir lediglich die Präzisionen austauschen. Hierfür lautet die analoge Formel

$$N_{ad,eff} = N_{hist} \cdot \frac{V_{ad,\tau,\mu_{shift}}^{-1}}{V_{ad,\tau=0,\mu_{shift_0}}^{-1}} = N_{hist} \cdot \omega_{ad}. \quad (3.8)$$

Wie wir an der Definition bereits erkennen können, steht die Precision Weight Effective Sample Size bereits vor Planung der neuen Studie fest. Unabhängig von den Planungsmodalitäten bleibt sie konstant. Trivialerweise haben aber die informativen a-priori Parameter im Modell der neuen Studie je nach Annahmen einen anderen Einfluss auf den Ausgang der Studie und somit auch auf die Power. So wie die Prior Effective Sample Size daher in *Neuenschwander et al.* [50] definiert wird, kann sie nicht mit der Anzahl der Patienten übereinstimmen, um die eine neue Studie ohne Powerverlust reduziert werden kann. Damit Formel (3.7) gilt, müsste die folgende lineare Beziehung zwischen der Präzision und der Fallzahl gelten

$$N_{hist} = c \cdot \frac{1}{V_{ped,\tau=0,\mu_{shift_0}}}. \quad (3.9)$$

mit einer unbekanntem Konstanten $c \in \mathbb{R}$. Dieser Zusammenhang ist jedoch nicht erfüllt.

Beispiel 3.1. Betrachten wir hierzu folgendes Beispiel. Wir nehmen an, es existiere nur ein Kollektiv und eine historische Studie h mit gleichgroßen Behandlungsgruppen. Die binäre Zielgröße sei das Log Odds Ratio

$$\theta_h = \log OR = \log \left(\frac{p_T / (1 - p_T)}{p_C / (1 - p_C)} \right), \quad (3.10)$$

mit den Ereignisraten p_T in der Behandlungsgruppe und p_C in der Kontrollgruppe. Die jeweils zugehörigen Fallzahlen seien n_T und n_C . Mit der δ -Methode gilt für den quadrierten Standardfehler [49, Formel (33), (35)]

$$\sigma_h^2 = \frac{1}{p_T(1-p_T) \cdot n_T} + \frac{1}{p_C(1-p_C) \cdot n_C} \quad (3.11)$$

Mit $N_{hist}/2 = n_T = n_C$ ergibt sich

$$\sigma_h^2 = \frac{1}{N_{hist}} \left(\frac{2}{p_T(1-p_T)} + \frac{2}{p_C(1-p_C)} \right) = \frac{1}{N_{hist}} \cdot s_h^2, \quad (3.12)$$

mit der empirischen Intrastudienvarianz s_h^2 . Das zugrundeliegende Modell in diesem Beispiel sei das hierarchische Normal-Normal-Modell aus Kapitel 2.4.2. Dabei sei τ^2 bekannt. Als hyper a-priori Verteilung für μ verwenden wir eine nicht-informative Gleichverteilung. In dieser Situation ist die a-posteriori prädiktive Verteilung von θ^* eine Normalverteilung der Gestalt (2.40). Da nur eine historische Studie existiert, reduziert sich die a-posteriori prädiktive Varianz V_τ auf die Summe der a-posteriori Varianz V_μ und der Interstudienvarianz (vgl. Formeln (2.7), (2.40)). Mit Gleichung (3.12) gilt

$$V_\tau = V_\mu + \tau^2 = \frac{1}{\frac{1}{\sigma_h^2} + \tau^2} + \tau^2 = \sigma_h^2 + 2\tau^2 = \frac{1}{N_{hist}} \cdot s_h^2 + 2\tau. \quad (3.13)$$

Da τ bei einer historischen Studie nicht geschätzt werden kann, aber im Allgemeinen nicht 0 ist, stellt die Präzision kein Vielfaches der historischen Fallzahl dar. Damit ist die Bedingung (3.9) nicht erfüllt. Ein alternatives Gegenbeispiel ist im Anhang angegeben (vgl. Kap. A.1).

Die Precision Weight Effective Sample Size ist folglich nicht genau die Anzahl Patienten, die durch die Extrapolation der historischen Ergebnisse, in einer neuen Studie eingespart werden kann. Da sie aber von *Neuenschwander et al.* [50] und *Gsteiger et al.* [30] in diesem Sinne interpretiert wird, ist dies für uns der Anlass, sie in der Simulationsstudie (Kap. 5) genauer zu untersuchen. Hier kontrollieren wir, ob die Power in der neuen Studie eingehalten wird, wenn wir die benötigte Fallzahl um N_{eff} Teilnehmer reduzieren und gleichzeitig das historische Vorwissen in die neue Studie einbringen. Im Verlauf dieser Arbeit werden wir die Precision Weight Effective Sample Sizes $N_{ped,eff}$ und $N_{ad,eff}$ als Maß für die jeweiligen Gewichtungsfaktoren ω_{ped} und ω_{ad} der Präzisionen auffassen. In den Anwendungen in Kapitel 4 werden wir darum den Einfluss der unterschiedlichen hyper a-priori Verteilungen auf die

Gewichtungsfaktoren anhand der Precision Weight Effective Sample Sizes veranschaulichen und diskutieren.

Wenn nur ein historisches Kollektiv existiert und die echte Interstudienvarianz τ^2 bekannt ist, können wir die Präzisionen in Gleichung (3.6) analytisch bestimmen. Somit können wir asymptotisch die „wahre“ Precision Weight Effective Sample Size bzw. den „wahren“ Gewichtungsfaktor effektiv berechnen. Diese bezeichnen wir mit N_{true} bzw. ω_{true} . Hierfür können wir auf die Erkenntnisse aus Abschnitt 2.4.2 zurückgreifen. Wir betrachten $h \in \{1, \dots, H\}$ frühere Studien an einem Patientenkollektiv. Es bezeichne V_τ die a-posteriori prädiktive Varianz von θ^* im vollen Modell und $V_{\tau \equiv 0}$ die im Referenzmodell. Unter der Annahme von bekanntem τ gilt für die Varianzen nach Formel (2.40) approximativ

$$V_\tau = \text{Var}(\theta^* | y_1, \dots, y_H, \tau) = \frac{1}{\sum_{h=1}^H \left(\frac{1}{\sigma_h^2 + \tau^2} \right) + \tau^2}, \quad (3.14)$$

$$V_{\tau \equiv 0} = \text{Var}(\theta^* | y_1, \dots, y_H, \tau \equiv 0) = \frac{1}{\sum_{h=1}^H \left(\frac{1}{\sigma_h^2} \right)}. \quad (3.15)$$

Setzen wir die Gleichungen (3.14) und (3.15) in Formel (3.7) ein, ergibt sich

$$\begin{aligned} N_{true} &:= N_{eff | \tau_{\text{bekannt}}} \\ &= \omega_{true} \cdot N_{hist} \\ &= \frac{\frac{1}{V_\tau}}{\frac{1}{V_{\tau \equiv 0}}} \cdot N_{hist} \\ &= \frac{V_{\tau \equiv 0}}{V_\tau} \cdot N_{hist} \\ &= \frac{1}{\sum_{h=1}^H \left(\frac{1}{\sigma_h^2} \right)} \cdot N_{hist} \\ &= \frac{1}{\sum_{h=1}^H \left(\frac{1}{\sigma_h^2 + \tau^2} \right) + \tau^2} \cdot N_{hist} \\ &= \frac{\left(\sum_{h=1}^H \sigma_h^{-2} \right)^{-1}}{\left(\sum_{h=1}^H (\sigma_h^2 + \tau^2)^{-1} \right)^{-1} + \tau^2} \cdot N_{hist}. \end{aligned} \quad (3.16)$$

3.3.3. Validierung und Poweranalyse

In diesem Abschnitt stellen wir die Planung und Durchführung einer neuen Studie unter Nutzung von Vorwissen vor. Ziel ist die Validierung der ermittelten Precision Weight Effective Sample Size $N_{ped,eff}$. Das Vorwissen aus den historischen Studien wird über die a-priori Verteilung in das Normal-Normal-Modell der neuen Studie eingebracht. In Kapitel 4 werden wir einige klinische Beispiele betrachten. Für diese ermitteln wir die a-posteriori prädiktive Verteilung von θ^* und berechnen die Precision Weight Effective Sample Size. Anschließend überprüfen wir, ob die virtuelle Anzahl Patienten tatsächlich der a-posteriori prädiktiven Verteilung entspricht. Hierzu benutzen wir die a-posteriori prädiktiven Parameter als informative a-priori Parameter in der neuen Studie. Wir rekrutieren N_{eff} Patienten weniger als ursprünglich ohne Vorwissen benötigt und überprüfen, ob die geplante Power eingehalten wird.

Im Heparin-Beispiel 4.1 und im FeNO-Beispiel 4.3 vergleichen wir eine neue Behandlung mit der Standardbehandlung bezüglich eines binären Endpunkts. In beiden Beispielen planen wir eine neue Studie an Kindern. Die neue Studie soll die Überlegenheit der innovativen Behandlung gegenüber der Standardtherapie bezüglich des binären Endpunkts zeigen. Beispielsweise kann dies bedeuten, dass unter der innovativen Therapie weniger unerwünschte Ereignisse eintreten sollen. Die einseitige Nullhypothese H_0 und die Alternativhypothese H_1 sind gegeben durch

$$H_0 : OR = \frac{Odds(p_T)}{Odds(p_C)} \geq 1 \quad \text{versus} \quad H_1 : OR = \frac{Odds(p_T)}{Odds(p_C)} < 1, \quad (3.17)$$

wobei p_T die Ereignisrate unter der innovativen Therapie und p_C die Ereignisrate unter der Standardbehandlung seien. Als erstes bestimmen wir die Fallzahl N_{ref} für die zukünftige Studie ohne Nutzung der historischen Vorinformation. Aus Gründen der Vergleichbarkeit mit der späteren Simulationsstudie bestimmen wir die Fallzahl zur Validierung der Precision Weight Effective Sample Size innerhalb eines bayesianischen Simulationsansatzes. Wir simulieren die Power, weil wir nach dem gleichen Prinzip später die historischen Studienergebnisse für die Metaanalysen generieren (vgl. Kap. 5.1.3). Wir werden aber sehen, dass wir für dieses spezielle Modell die Power auch analytisch berechnen können (vgl. Kap. 3.4). Die Fallzahlschätzung stimmt aufgrund der simulierten Power folglich nicht exakt mit der Planung im frequentistischen Fall bzw. bei analytischer Berechnung der Power überein [vgl. 62, S. 157].

Für die Überprüfung der Hypothesen betrachten wir erneut ein Normal-Normal-Modell mit nicht-informativer a-priori Verteilung. Dadurch gewährleisten wir, dass die a-posteriori Verteilung durch die Beobachtungen der neuen Studie bestimmt wird. Das Modell für die neue Studie lautet:

Datenmodell:

$$y|\theta \sim N(\theta, \sigma^2) \quad \text{mit beobachteten } \textit{LogOR} \textit{ in der neuen Kinderstudie } y, \\ \text{zugehörigem studienspezifischen Effekt } \theta \\ \text{und fester Intrastudienvarianz } \sigma^2.$$

Parametermodell:

$$\theta|\mu_0, \tau_0^2 \sim N(\mu_0, \tau_0^2) \quad \text{mit festem Erwartungswert } \mathbb{E}(\theta|\mu_0, \tau_0^2) = \mu_0 = 0 \\ \text{und fester Varianz } \mathbb{V}\text{ar}(\theta|\mu_0, \tau_0^2) = \tau_0^2 = \infty. \quad (3.18)$$

In diesem Design können wir die a-posteriori Verteilung von θ direkt berechnen. Da τ_0^2 fest ist und lediglich eine Studienbeobachtung, das beobachtete Log Odds Ratio, existiert, befinden wir uns in der Situation des Normal-Modells bei einer Beobachtung (Kap. 2.2.1) bzw. der des Modells 2.4.2. Die a-posteriori Verteilung ergibt sich daher unmittelbar durch Einsetzen der gewählten Hyperparameter in Formel (2.7) oder Formel (2.34) und (2.35)

$$\theta|y \sim N(\mu_1, \tau_1^2) \quad \text{mit} \quad \mu_1 = \frac{\frac{1}{\tau_0^2}\mu_0 + \frac{1}{\sigma^2}y}{\frac{1}{\tau_0^2} + \frac{1}{\sigma^2}} \stackrel{(3.18)}{=} \frac{\frac{1}{\sigma^2}y}{\frac{1}{\sigma^2}} = y \\ \text{und} \quad \tau_1^2 = \frac{1}{\frac{1}{\tau_0^2} + \frac{1}{\sigma^2}} \stackrel{(3.18)}{=} \frac{1}{\sigma^{-2}} = \sigma^2. \quad (3.19)$$

Somit ist $\theta|y \sim N(y, \sigma^2)$ verteilt mit Stichprobenvarianz σ^2 . In der klassischen Statistik würden wir zur Überprüfung der Nullhypothese einen einseitigen Signifikanztest zu einem festgelegten Niveau α , zum Beispiel $\alpha = 2.5\%$ durchführen. Entsprechend definieren für diese bayesianische Analyse „signifikant“, wenn die Wahrscheinlichkeit, dass der a-posteriori Parameter θ kleiner 0 ist, mehr als $1 - \alpha = 97.5\%$ beträgt, also

$$\mathbb{P}(\theta < 0|y) \geq 97.5\%. \quad (3.20)$$

Das Ziel der Fallzahlschätzung ist eine bayesianische Power von 80%

$$\text{Power} = \mathbb{P}_{p_C, p_T}(\mathbb{P}(\theta < 0|y) \geq 97.5\%) = 0.8. \quad (3.21)$$

Die angenommenen festen Ereignisraten p_T und p_C in der zukünftigen Kinderstudie setzen wir in einigen Beispielen jeweils gleich den Ereignisraten in der historischen Kinderstudie bzw. gleich den gepoolten Ereignisraten der Kinderstudien. Für die neue Erwachsenenstudie nehmen wir das historische prädiktive Log Odds Ratio und die gemittelte Ereignisrate beider Behandlungsgruppen als Planungsgrundlage an, so dass sich die Ereignisraten p_T und p_C direkt daraus ergeben. Um die Power bei vorgegebener Fallzahl zu bestimmen, simulieren wir 1 000 000 mal das Ergebnis einer zukünftigen Studie mit den vorausgesetzten Ereignisraten. Dafür generieren wir 1 000 000 binomialverteilte Ziehungen

$$e_T \sim \text{Bin}(n_T, p_T),$$

$$e_C \sim \text{Bin}(n_C, p_C),$$

wobei e_T die Anzahl der Ereignisse unter der neuen Therapie und e_C die Anzahl der Ereignisse unter der Standardtherapie bezeichnet. Dabei sind n_T und n_C die Patientenzahlen im jeweiligen Behandlungsarm. Wir nehmen an, dass die Gruppen gleich groß sind, also $n_C = n_T$. Für jede Ziehung berechnen wir das zugehörige beobachtete Log Odds Ratio θ_{ob} mit beobachteter Intrastudienvarianz σ_{ob}^2

$$\theta_{ob} = \log\left(\frac{e_T(n_C - e_C)}{e_C(n_T - e_T)}\right) \quad \text{und} \quad \sigma_{ob}^2 = \frac{1}{e_T} + \frac{1}{n_T - e_T} + \frac{1}{e_C} + \frac{1}{n_C - e_C}. \quad (3.22)$$

Wir bestimmen die benötigte Referenzfallzahl $N_{ref} := n_C + n_T$ ohne Einbringung von Vorwissen so, dass für 80% der simulierten Studien die obere Grenze des einseitigen 97.5% Credible- bzw. Konfidenzintervalls von θ kleiner als 0 ist. Da die a-posteriori Verteilung normalverteilt ist, können wir die obere Grenze des Konfidenzintervalls wie folgt bestimmen

$$\text{UCL}_{0.975} = \theta_{ob} + z_{0.975} \sqrt{\sigma_{ob}^2}. \quad (3.23)$$

Hier bezeichne $z_{0.975}$ das 97.5% Quantil der Standardnormalverteilung. Im Anschluss überprüfen wir, in wie vielen Simulationen die obere Schranke kleiner als 0 ist. Um die benötigte Fallzahl N_{ref} zu bestimmen, variieren wir N_{ref} so lange, bis in 80% der Ziehungen diese Bedingung erfüllt ist.

Im nächsten Schritt analysieren wir, ob die bayesianische Power unter Einbringung des extrapolierten Vorwissens eingehalten wird. Dazu modifizieren wir das Normal-Normal-Modell (3.18) durch informative Hyperparameter. Für den festen Hyperparameter μ_0 wählen wir genau den a-posteriori prädiktiven Erwartungswert, den wir im vollen Modell 3.2 bestimmt

haben. Für τ_0^2 wählen wir analog die ermittelte a-posteriori prädiktive Varianz aus dem vollen Modell. Die Parameterebene des informativen Modells ist somit

Parametermodell:

$$\theta | \mu_0, \tau_0^2 \sim N(\mu_0, \tau_0^2)$$

$$\text{mit festem Erwartungswert} \quad \mathbb{E}(\theta | \mu_0, \tau_0^2) = \mu_0 \equiv \mathbb{E}(\theta_{ped}^* | y_1, \dots, y_{A+H}) \quad (3.24)$$

$$\text{und fester Varianz} \quad \mathbb{V}\text{ar}(\theta | \mu_0, \tau_0^2) = \tau_0^2 \equiv \mathbb{V}\text{ar}(\theta_{ped}^* | y_1, \dots, y_{A+H}) = V_{ped, \tau, \mu_{shift}}.$$

In der neuen Studie reduzieren wir die Fallzahl N_{ref} um N_{eff} und rekrutieren dementsprechend weniger Patienten. Wir simulieren 100 000 mal erneut die möglichen Ergebnisse einer neuen Studie und überprüfen, ob das einseitige Konfidenzintervall die Null nicht überdeckt. Im Anschluss kontrollieren wir, inwieweit die ermittelte Power von der geplanten Power ohne informative a-priori Verteilung abweicht. Falls die berechnete Precision Weight Effective Sample Size N_{eff} der a-posteriori prädiktiven Verteilung von θ^* entspricht, sollte die Power trotz reduzierter Fallzahl ebenfalls 80% betragen.

3.4 Prior Effective Sample Size

Für eine spezielle neue Studie können wir auf eine alternative Weise die tatsächlich Fallzahl berechnen, die unter Berücksichtigung des Vorwissens eingespart werden kann. Wir bezeichnen diese Fallzahl als *Prior Effective Sample Size* N_{prior} . Wir können sie berechnen, indem wir informative Prior benutzen und die Fallzahl der neuen Studie, die unter Verwendung nicht-informativer a-priori Verteilungen benötigt wird, so lange reduzieren, bis die Power das geplante Niveau erreicht. Diese Anzahl spiegelt exakt den Informationswert der informativen a-priori Verteilung für diese eine neue Studie wieder. Damit trägt die Fallzahl ihren Namen zurecht. Die allgemeine Definition, die für ein beliebiges Modell zur Überprüfung einer einseitigen Nullhypothese $H_0 : \theta \geq c$ mit Signifikanzniveau α und Power $1 - \beta$ gilt, lautet:

Definition 3.1. *Die einseitige Nullhypothese der neuen Studie sei $H_0 : \theta \geq c$ mit $c \in \mathbb{R}$. Es sei $N_{ref} \in \mathbb{N}$ die Fallzahl, die zur Überprüfung von H_0 mit Signifikanzniveau α , Power $1 - \beta$ und nicht-informativer a-priori Verteilungen benötigt wird. Die zugehörige a-posteriori Verteilung von θ sei $p_{uninf}(\theta | y)$ sowie \mathbb{P}_{uninf} die Wahrscheinlichkeit eines Events. Ebenso sei $N_{ref, inf} \in \mathbb{N}$ die benötigte Fallzahl unter Anwendung informativer a-priori Verteilungen,*

$p_{inf}(\theta|y)$ die entsprechende a-posteriori Verteilung und \mathbb{P}_{inf} die Wahrscheinlichkeit. Dann ist die Prior Effective Sample Size (PESS) definiert als

$$N_{prior} := N_{ref} - N_{ref,inf} \quad (3.25)$$

$$\text{mit } N_{ref} = \inf \{n \in \mathbb{N} : \mathbb{P}(\mathbb{P}_{uninf}(\theta < c | y) \geq 1 - \alpha) \geq 1 - \beta\}$$

$$\text{mit } N_{ref,inf} = \inf \{n \in \mathbb{N} : \mathbb{P}(\mathbb{P}_{inf}(\theta < c | y) \geq 1 - \alpha) \geq 1 - \beta\}.$$

Für den in Kapitel 3.3.3 vorgestellten Spezialfall können wir die Formel für die Power analytisch herleiten und damit auch direkt N_{prior} berechnen.

Beispiel 3.2. In der neuen Studie soll die einseitige Fragestellung (3.17) mit Signifikanzniveau $\alpha = 0.025$ und Power=80% mit dem Normal-Modell bei einer Beobachtung und bekannter Varianz σ^2 betrachtet werden. Die Beobachtung y ist das Log Odds Ratio der neuen Studie. Für die a-posteriori Verteilung gilt somit $\theta|y \sim N(\mu_1, \tau_1^2)$ (2.6) mit den Parametern in Gleichung (2.7). Aufgrund der normalverteilten a-posteriori Verteilung lautet die Regel für die Testentscheidung

$$\begin{aligned} \text{Nullhypothese } H_0 \text{ ablehnen} &\Leftrightarrow \mathbb{P}(\theta < 0|y) \geq 0.975 \\ &\Leftrightarrow \mathbb{P}\left(\frac{\theta - \mu_1}{\tau_1} \leq -\frac{\mu_1}{\tau_1} \middle| y\right) \geq 0.975 \\ &\Leftrightarrow \Phi\left(-\frac{\mu_1}{\tau_1}\right) \geq 0.975, \end{aligned} \quad (3.26)$$

mit der Verteilungsfunktion Φ der Standardnormalverteilung. Damit ist die Power für einen festen Parameter θ und gegebene Fallzahl n

$$\begin{aligned} \text{Power} &= \mathbb{P}(\mathbb{P}(\theta < 0|y) \geq 0.975) \\ &= \mathbb{P}\left(\Phi\left(-\frac{\mu_1}{\tau_1}\right) \geq 0.975\right) \\ &= \mathbb{P}\left(-\frac{\mu_1}{\tau_1} \geq z_{0.975}\right) \\ &= \mathbb{P}\left(\frac{\mu_1}{\tau_1} \leq z_{0.025}\right), \end{aligned} \quad (3.27)$$

mit dem 97.5%-Quantil und dem 2.5%-Quantil der Standardnormalverteilung. Wie im Modell 3.19 gezeigt, gilt für die Parameter der a-posteriori Normalverteilung unter Verwendung nicht-

informativer Prior: $\mu_1 = y$ und $\tau_1^2 = \sigma^2$. Da außerdem für die Beobachtung $y|\theta \sim N(\theta, \sigma^2)$ gilt, folgt durch Einsetzen in Formel (3.27) für die *Power ohne Berücksichtigung von Vorwissen*

$$\begin{aligned} \text{Power}_{uninf} &= \mathbb{P}_{uninf} \left(\frac{y}{\sigma} \leq z_{0.025} \right) \\ &= \mathbb{P}_{uninf} \left(\frac{y - \theta}{\sigma} \leq z_{0.025} - \frac{\theta}{\sigma} \right) \\ &= \Phi \left(z_{0.025} - \frac{\theta}{\sigma} \right). \end{aligned} \quad (3.28)$$

Die Power unter Anwendung einer informativen a-priori Verteilung können wir in diesem konkreten Fall ebenfalls berechnen. Da nur eine Studie bzw. Beobachtung existiert, kennen wir die allgemeine Gestalt des a-posteriori Erwartungswertes μ_1 und der a-posteriori Varianz τ_1^2 im Normal-Modell (vgl. Gleichung (2.7)). Setzen wir diese in Formel (3.27) ein, so ergibt sich für die *Power unter Benutzung informativer a-priori Verteilungen*

$$\begin{aligned} \text{Power}_{inf} &= \mathbb{P}_{inf} \left(\mu_1 \cdot \tau_1^{-1} \leq z_{0.025} \right) \\ &= \mathbb{P}_{inf} \left(\frac{\frac{1}{\tau_0^2} \mu_0 + \frac{1}{\sigma^2} y}{\frac{1}{\tau_0^2} + \frac{1}{\sigma^2}} \cdot \sqrt{\frac{1}{\tau_0^2} + \frac{1}{\sigma^2}} \leq z_{0.025} \right) \\ &= \mathbb{P}_{inf} \left(\frac{\frac{1}{\tau_0^2} \mu_0 + \frac{1}{\sigma^2} y}{\sqrt{\frac{1}{\tau_0^2} + \frac{1}{\sigma^2}}} \leq z_{0.025} \right) \\ &= \mathbb{P}_{inf} \left(\frac{1}{\tau_0^2} \mu_0 + \frac{1}{\sigma^2} y \leq z_{0.025} \sqrt{\frac{1}{\tau_0^2} + \frac{1}{\sigma^2}} \right) \\ &= \mathbb{P}_{inf} \left(y \leq \left(z_{0.025} \sqrt{\frac{1}{\tau_0^2} + \frac{1}{\sigma^2}} - \frac{1}{\tau_0^2} \mu_0 \right) \cdot \sigma^2 \right) \\ &= \mathbb{P}_{inf} \left(\frac{y - \theta}{\sigma} \leq \left[\frac{\left(z_{0.025} \sqrt{\frac{1}{\tau_0^2} + \frac{1}{\sigma^2}} - \frac{1}{\tau_0^2} \mu_0 \right) \cdot \sigma^2 - \theta}{\sigma} \right] \right) \\ &= \Phi \left(\frac{1}{\sigma} \left[\left(z_{0.025} \sqrt{\frac{1}{\tau_0^2} + \frac{1}{\sigma^2}} - \frac{1}{\tau_0^2} \mu_0 \right) \cdot \sigma^2 - \theta \right] \right). \end{aligned} \quad (3.29)$$

Die letzte Gleichung gilt, da $y|\theta \sim N(\theta, \sigma^2)$ verteilt ist. Somit können wir in diesem Modell die Power mit und ohne Nutzung von Vorwissen explizit berechnen. In komplizierteren Modellen kann die Power entsprechend simuliert werden. Mit Formel (3.28) und (3.29) lautet die Definition der Prior Effective Sample Size N_{prior} für dieses Modell

$$N_{prior} := N_{ref} - N_{ref,inf} \quad (3.30)$$

$$\text{mit } N_{ref} = \inf \left\{ n \in \mathbb{N} : \Phi \left(z_{0.025} - \frac{\theta}{\sigma} \right) \geq 0.8 \right\}$$

$$\text{und } N_{ref,inf} = \inf \left\{ n \in \mathbb{N} : \Phi \left(\frac{1}{\sigma} \left[\left(z_{0.025} \sqrt{\tau_0^{-2} + \sigma^{-2}} - \tau_0^{-2} \mu_0 \right) \cdot \sigma^2 - \theta \right] \right) \geq 0.8 \right\}.$$

In den klinischen Beispielen 4.1 und 4.3 werden wir zusätzlich zur Analyse in Kapitel 3.3.3 die Fallzahl der neuen Studie erneut unter Anwendung der Powerformel (3.28) bestimmen. Anschließend berechnen wir die Prior Effective Sample Size N_{prior} bei Anwendung der a-posteriori prädiktiven Parameter des vollen Modells als informative a-priori Parameter der neuen Studie.

3.5 MCMC-Einstellungen

Für die Bestimmung der verschiedenen a-posteriori Verteilungen greifen wir auf Markov Chain Monte Carlo Verfahren zurück (vgl. Kap. 2.4.1). In allen Modellen generieren wir jeweils 2 Markovketten. Jede besteht aus 110 000 Updating-Schritten. Nachdem wir die Konvergenz in allen Anwendungsbeispielen überprüft haben, wählen wir für unsere Anwendungen standardmäßig 10 000 Iterationen als Burn-In. Der Burn-In ist die Anzahl an Simulationsschritten, die die Markovkette benötigt, um sich „einzupendeln“ bzw. um die stationäre Verteilung zu erreichen. Die Burn-In Schritte werden für die Schätzung der Parameter nicht berücksichtigt. In unseren Anwendungsbeispielen gehen wir davon aus, dass die Markovkette nach dem Burn-In von 10 000 bereits gegen die stationäre Verteilung konvergiert. Desweiteren verwenden wir für die drei Anwendungsbeispiele in Kapitel 4 kein Thinning, während wir in den Simulationsszenarien in Kapitel 5 Thinning mit dem Wert 10 benutzen. Das bedeutet, dass nur jeder zehnte Iterationsschritt der Markovkette gespeichert wird. Die Markovkette wird, wie der Begriff schon sagt, ausgedünnt. Die einzelnen Iterationsschritte sind bedingte Ziehungen aus der *Full Conditional* Verteilung (siehe 2.21). Darum ist in unserem Zusammenhang das Ziel des Thinnings, die Autokorrelation zwischen den abhängigen Iterationsschritten zu senken. In den Simulationsszenarien macht dies Sinn, da dort nur wenige historische Beobachtungen bzw. Studien vorhanden sind. Dort müssen wir davon auszugehen, dass die nah beieinander liegenden Iterationen stärker miteinander korreliert sind, weil die gewählten a-priori Verteilungen einen höheren Einfluss auf die a-posteriori Werte haben. Der Sinn und die Notwendigkeit von Thinning werden teilweise kontrovers diskutiert [41]. Allerdings gibt es keine eindeutige Empfehlung gegen oder für die Anwendung von Thinning [41, S. 115].

Insgesamt erhalten wir nach Abzug des Burn-In und Berücksichtigung des Thinnings in den Anwendungsbeispielen jeweils $2 \cdot 100\,000$ und in den Simulationsszenarien jeweils $2 \cdot 10\,000$ auswertbare Iterationen zur Schätzung der a-posteriori Verteilungen.

3.6 Getrimmtes Modell

Das hier vorgestellte getrimmte Modell soll ausschließlich der Validierung der geschätzten Parameter dienen. Um herauszufinden wie stabil die Ergebnisse des vollen Modells 3.2 sind, wenden wir eine Trimming-Methode an. In der deskriptiven Statistik ist das getrimmte Mittel eine robuste Alternative zum Mittelwert bei normalverteilten Daten [4, 35]. Außerdem nutzt das α -getrimmte Mittel mehr Informationen der Stichprobe als der Median. Einzelne Extremwerte können den Mittelwert verfälschen und gleichzeitig die zugehörige Varianz extrem vergrößern. Die zugrundeliegende Idee ist, dass durch das Entfernen eines bestimmten Prozentsatzes α von minimalen und maximalen Beobachtungen aus der Stichprobe, die Schätzung stabiler wird. Das α -getrimmte Mittel ist dann der Mittelwert über die verbleibenden Beobachtungen. Durch das Entfernen von gleich vielen, symmetrisch verteilten, minimalen und maximalen Werten wird einzig die Streuung verkleinert. Die mittleren Lageparameter bleiben unbeeinflusst. Im Falle symmetrischer Verteilungen ist das α -getrimmte Mittel ein unverfälschter Schätzer für den Erwartungswert und den Median. Es ist asymptotisch normalverteilt [35, S. 1541].

Wir übertragen diese Idee in etwas abgewandelter Form auf das volle Normal-Normal-Modell (Kap. 3.2). Für unsere Anwendung ist θ^* der Parameter von Interesse und insbesondere die a-posteriori prädiktive Verteilung von θ^* . Für die Bestimmung dieser Verteilung verwenden wir MCMC-Verfahren mit den bereits erwähnten Voreinstellungen (Kap. 3.5). Nach der Berechnung des vollen Modells erhalten wir 200 000 bzw. 20 000 auswertbare Iterationen der Markovketten. Wir setzen α gleich 0.001. Anschließend beschneiden wir die simulierten Ketten bezüglich des jeweiligen θ^* . Konkret bedeutet dies, dass wir als erstes die Iterationen identifizieren, die die 0.1% minimalsten und 0.1% maximalsten Werte für θ^* beinhalten. Das sind für $\alpha = 0.001$ in absoluten Zahlen 400 von 200 000 bzw. 40 von 20 000 Iterationsschritten. Anschließend entfernen wir für alle zu schätzenden Parameter genau diese ermittelten Iterationen. Wir führen das Verfahren für alle Parameter durch, da wir annehmen, dass beispielsweise θ_{ped}^* dann extrem ist, falls gleichzeitig die Interstudienvarianz τ^2 groß wird. Konsequenterwei-

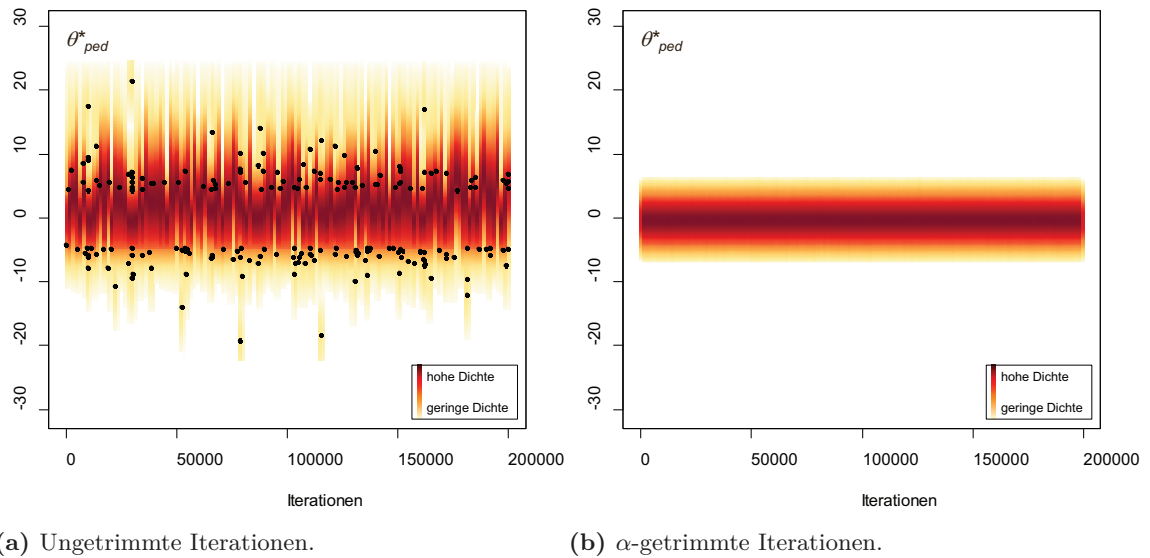


Abb. 3.2: Smooth-Scatterplots der α -Trimming-Methode.

Gepoolte Iterationsschritte der generierten Markovketten im FeNO-Beispiel (Kap. 4.3) mit hyper a-priori Verteilung $\tau^2 \sim \Gamma^{-1}(0.001, 0.001)$. Die schwarzen Punkte repräsentieren die Extrema, die entfernt werden. (MCMC-Simulationen: Chains=2, Iterationen=110 000, Burn-In=10 000, gültige Iterationen=200 000. Getrimmtes Modell: $\alpha = 0.001$, gültige Iterationen=199 600).

se nimmt die Variation von θ_{ped}^* durch das Entfernen der extremen Iterationen ab. Somit ist $V_{ped, \tau, \mu_{shift}}$ im getrimmten Modell kleiner als im vollen Modell. Der Gewichtungsfaktor der Präzisionen ω_{trim} (vgl. Gleichung 3.6) wird dadurch im getrimmten Modell konsequenter größer. Die historischen Ergebnisse erhalten mehr Gewicht. Folglich steigen auch die Precision Weight Effective Sample Size N_{eff} und die Prior Effective Sample Size N_{prior} . In Abbildung 3.2 sind die generierten Werte für das a-posteriori prädiktive Log Odds Ratio θ_{ped}^* zweier Markovketten im späteren Beispiel 4.3 dargestellt. Auf der linken Seite sind die ungetrimmten Simulationen dargestellt, wobei die Punkte genau die Extremwerte repräsentieren, die durch das Trimmen entfernt werden. Wir sehen, dass die Punkte sehr unrealistischen Werten für ein Log Odds Ratio entsprechen. Auf der rechten Seite sind die bereinigten Schritte der Markovkette abgebildet.

Wegen dieser Eigenschaften können wir das getrimmte Modell als Hilfsmittel benutzen, um die Stabilität der geschätzten Parameter des vollen Modells zu überprüfen. Unterscheidet sich beispielsweise die Precision Weight Effective Sample Size N_{eff} des getrimmten Modells nur sehr wenig von der des vollen Modells, ist dies ein Indiz dafür, dass die Schätzung von θ^* robust ist. Im Gegensatz dazu entspricht eine große Abweichung eher einer unsicheren Schätzung. Der Informationsgehalt der historischen Ergebnisse reicht dann nicht aus, um eine präzise Aussage über den studienspezifischen Effekt einer zukünftigen Studie zuzulassen.

Dieses α -getrimmte Modell ist nicht geeignet, um mit dessen Ergebnissen später tatsächlich eine klinische Studie zu planen und durchzuführen. Der Grund dafür ist, dass wir „uneingeschränkte“ a-priori-Verteilungen benutzen und von der a-posteriori Verteilung eine MCMC-Stichprobe ziehen. Erst im Nachhinein sehen wir, dass die Stichprobe unpassende Werte enthält und ändern ad-hoc die a-posteriori Verteilung. Der Kritikpunkt an diesem Vorgehen ist, dass wir dadurch datengesteuert indirekt auch die a-priori Verteilung ändern. Dies ist vergleichbar damit, zu einer bestimmten Fragestellung zwei unterschiedliche Signifikanztests durchzuführen und zur Überprüfung der Hypothese den Test zu wählen, welcher den kleineren p-Wert liefert. Ein gültiges Vorgehen bestünde darin, von vornherein abgeschnittene a-priori Verteilungen zu verwenden und dadurch Werte, die oberhalb bzw. unterhalb einer Grenze liegen, nicht zuzulassen. Der erlaubte Wertebereich sollte möglichst alle in der Praxis realistischen Werte der jeweiligen Zielgröße enthalten.

3.7 Software

Für die statistischen Analysen und Simulationen benutzen wir die Softwarepakete:

- R, R Foundation for Statistical Computing, Version 2.15.2 [vgl. 53],
- OpenBUGS, Version 3.2.2 rev 1063 [vgl. 44] ¹,
- WinBUGS, Version 1.4.3 [vgl. 43],
- SAS software, Version 9.3 of the SAS System for Windows ²,
- IBM SPSS Statistics for Windows, Version 21 ³.

Neben den bereits aufgeführten Statistikprogrammen verwenden wir zusätzliche R-Pakete. Das Paket R2WinBUGS [63] erlaubt ein direktes Interface zwischen R und OpenBUGS bzw. R und WinBUGS. Das Paket gtools [72] gibt uns die Möglichkeit Makroprogramme in R zu schreiben. Weitere Zusatzpakete, auf die wir zurückgreifen, sind standardmäßig in R implementiert bzw. werden innerhalb der anderen Pakete geladen.

¹ ©2010-2011 Members of OpenBUGS Project Management Group, 1990-1996 Medical Research Council (MRC), UK, 1996-2010 MRC, UK, and Imperial College, UK

² Copyright ©2012 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA

³ Copyright ©2012 IBM Corp. IBM SPSS Statistics for Windows, Version 21.0. Armonk, NY: IBM Corp.

III Ergebnisse

4 Anwendungen

Im Folgenden betrachten wir drei Anwendungsbeispiele der meta-analytisch prädiktiven Methode an reellen klinischen Studien. Hierbei zeigen wir in zwei Situationen mit vielen Erwachsenenstudien und nur einer Kinderstudie, wie unterschiedlich die Bewertung der Qualität der historischen Ergebnisse ausfallen kann. Im ersten Beispiel betrachten wir vergleichsweise homogene historische Studien mit einer binären Zielgröße. Im zweiten Beispiel verwenden wir sehr heterogene Studien mit einer stetigen Zielgröße. Die dritte Anwendung zeigt die Performance des meta-analytisch prädiktiven Ansatzes in der Situation weniger historischer Vorstudien. Dieses Beispiel ist die Motivation zur genaueren Untersuchung der Methodik in verschiedenen Simulationsszenarien (siehe Kap. 5).

4.1 Heparin-Studien

Studienbeschreibung. Als erste Anwendung der meta-analytisch prädiktiven Methode betrachten wir klinische Studien, die den Einsatz von Heparin an Patienten mit venösen Thromboembolien untersuchen. Das Glykosaminoglykan Heparin löst die Bindung von Blutgerinnungsfaktoren aus und wirkt somit hemmend auf die Gerinnungskaskade im Blut [42, S. 195f.]. Als Datengrundlage der Untersuchung an Erwachsenen dient der Cochrane Review „*Fixed dose subcutaneous low molecular weight heparins versus adjusted dose unfractionated heparin for venous thromboembolism*“ von *van Dongen et al.* [68]. Die interventionelle Behandlung, eine subkutane Injektion einer festen Dosis niedermolekularen (fraktionierten) Heparins, wird mit der konventionellen Behandlung, einer adjustierten Dosis unfraktionierten Heparins, verglichen. Niedermolekulares Heparin soll eine längere Wirkdauer nach subkutaner Injektion besitzen und somit die Applikationsintervalle vergrößern. Die Wirksamkeit der Therapien wurde auf einen binären Endpunkt reduziert. Dieser beschreibt das Auftreten von rekurrenten venösen Tromboembolien bis zum Ende der Nachbeobachtungszeit. Für den Review wurden die Ergebnisse von insgesamt $A = 18$ klinischen Studien mit 8122 erwachsenen Patienten zu dieser Fragestellung betrachtet. Insgesamt ergibt sich als Effektgröße ein gepooltes Odds Ratio von 0.68 [95% CI 0.55, 0.87]. Die Outcomes der einzelnen Studien sind in Abbildung 4.1 dargestellt. Wir haben den gepoolten Effekt mittels eines frequentistischen Random-Effects-

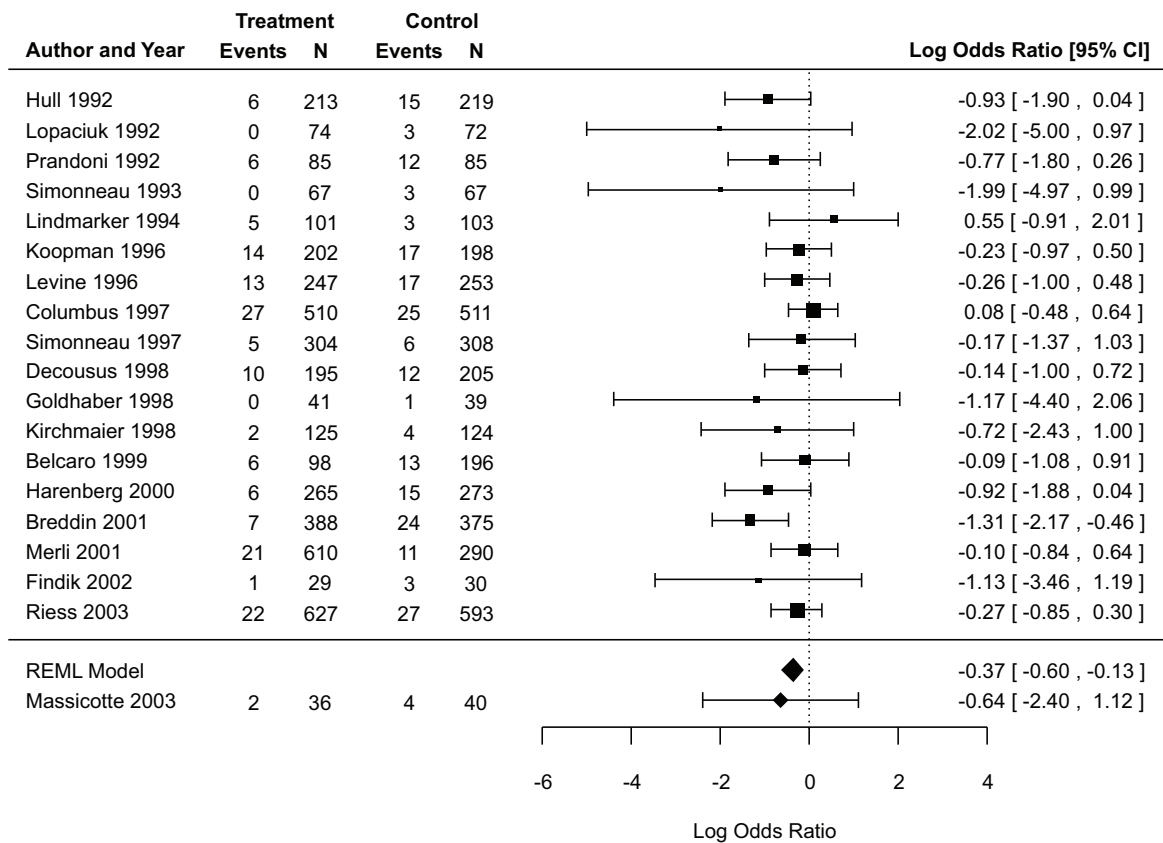


Abb. 4.1: Forestplot der Heparin-Studien.

Heterogenität (Erwachsenenstudien): $\tau^2 = 0.0176$ (95% CI 0, 0.3321), $df=17$ ($p=0.1325$), $I^2 = 6.89\%$ (95% CI 0, 58.33%). Test für Gesamteffekt (Erwachsenenstudien): $p=0.002$.

Modells (Restricted Maximum-Likelihood (REML) [51, 70]) berechnet. Die Annahme eines Zufallseffekts ist konsistenter zu unserer verwendeten bayesianischen Methodik als das im Artikel [68] verwendete Peto-Odds Ratio für feste Effekte. Obwohl es Unterschiede in der Zielgröße zwischen den Studien gibt, ist die ermittelte Heterogenität mit $I^2 = 6.89\%$ sehr gering. Die Intrastudienvarianz in den einzelnen Studien unterscheidet sich, was aber vor allem der unterschiedlichen Fallzahl der einzelnen Studien geschuldet ist. Im Gegensatz dazu ist die geschätzte Interstudienvarianz $\tau^2 = 0.0176$ sehr niedrig. Entsprechend können wir die Erwachsenenstudien als homogen erachten.

Neben den Untersuchungen an Erwachsenen gibt es zu dieser Fragestellung nur eine Kinderstudie mit vergleichbarem Behandlungskonzept, Outcome und vergleichbaren Einschlusskriterien. *Massicotte et al.* [47] berichten als Behandlungsergebnis an 76 Kindern ein Odds Ratio von 0.53 [95% CI 0.09; 3.08]. Ein Vergleich des *LogOR* der einzelnen Kinderstudie mit dem gepoolten *LogOR* der Erwachsenen zeigt einen tendenziell größeren Therapieeffekt in der Kinderpopulation. Allerdings ist dieser Trend aufgrund des Fehlens weiterer Kinderstudien

mit Vorsicht zu interpretieren. Diese Unsicherheit wird im vollen hierarchischen Modell in der a-posteriori Verteilung des Verschiebungsparameters μ_{shift} berücksichtigt.

Die Information der historischen $N_{hist} = 8198$ Patienten benutzen wir für die Planung und Auswertung einer neuen Studie an Kindern. Ziel der zukünftigen Studie ist der Nachweis der Überlegenheit der neuen Therapie gegenüber der Standardtherapie. Es gilt $OR < 1$ zu zeigen. Andererseits können wir die historischen Resultate ebenfalls zur Planung einer Studie an einem Patientenkollektiv von Erwachsenen nutzen. Da nur wenig über die Wirksamkeit der Behandlung an Kindern bekannt, aber viele Studien an Erwachsenen vorliegen, ist es klinisch sinnvoller, die historischen Daten für die Planung einer neuen Kinderstudie zu nutzen. Deshalb werden wir uns vorrangig auf die klinisch relevantere Extrapolation der historischen Daten für eine neue Kinderstudie beschränken und für eine neue Erwachsenenstudie nur die Ergebnisse zeigen.

Exemplarische Berechnung von ω_{ped} und $N_{ped,eff}$. Am Beispiel der Half-Normal-Verteilung als hyper a-priori Verteilung für den Varianzparameter τ zwischen den Studien, berechnen wir den Gewichtungsfaktor ω_{ped} und die entsprechende Fallzahl $N_{ped,eff}$ im Detail. Für alle anderen hyper a-priori Verteilungen und weiteren Anwendungsbeispiele folgt die Bestimmung analog. Darum geben wir für diese nur noch die Ergebnisse des meta-analytisch prädiktiven Ansatzes an.

Der erste Schritt besteht in der Kalkulation des vollen Modells (vgl. 3.2). Dafür greifen wir auf MCMC-Verfahren, wie sie in den Softwarepaketen OpenBUGS [44] und R [53] integriert sind, zurück. Wir betrachten, wie in Kapitel 3.5 beschrieben, zwei Markovketten mit jeweils 110 000 Iterationen und einem Burn-In von 10 000. Im vollen Modell schätzen wir die a-posteriori prädiktiven Verteilungen für das Log Odds Ratio einer neuen Kinderstudie θ_{ped}^* und für das Log Odds Ratio einer neuen Erwachsenenstudie θ_{ad}^* . Zusätzlich erhalten wir die a-posteriori Verteilungen für die Standardabweichung zwischen den Studien τ sowie für den Verschiebungsparameter μ_{shift} . Die a-posteriori Erwartungswerte und Varianzen der Parameter aus dem vollen Modell lauten

$$\begin{aligned} \mathbb{E}(\theta_{ad}^*|y) &= -0.3862, & V_{ad,\tau,\mu_{shift}} &= 0.3063^2, \\ \mathbb{E}(\theta_{ped}^*|y) &= -0.6145, & V_{ped,\tau,\mu_{shift}} &= 0.9542^2, \\ \mathbb{E}(\mu_{shift}|y) &= -0.2289, & \text{Var}(\mu_{shift}|y) &= 0.9271, \\ & & \mathbb{E}(\tau|y) &= 0.2262. \end{aligned}$$

Als Nächstes berechnen wir das Referenzmodell (Kap. 3.3.1). Wir benutzen den a-posteriori Erwartungswert für μ_{shift} aus dem vollen Modell als festen Hyperparameter im Referenzmodell $\mu_{shift_0} = -0.2289$. Zusätzlich nehmen wir an, dass die Interstudienvarianz identisch Null ist und setzen den Hyperparameter $\tau_0^2 \equiv 0$. Damit erhalten wir unter den gleichen MCMC-Einstellungen (MC=2, Iterationen=110 000, Burn-In=10 000) die Referenzvarianz

$$V_{ped, \tau \equiv 0, \mu_{shift_0}} = 0.1120^2.$$

Mit Formel (3.7) können wir die Precision Weight Effective Sample Size $N_{ped,eff}$ für die neue Kinderstudie bestimmen

$$\begin{aligned} N_{ped,eff} &= \omega_{ped} \cdot N_{hist} \\ &= \frac{1}{V_{ped, \tau, \mu_{shift}}} \cdot N_{hist} \\ &= \frac{1}{\frac{1}{V_{ped, \tau \equiv 0, \mu_{shift_0}}}} \cdot N_{hist} \\ &= \frac{1}{\frac{0.9542^2}{1}} \cdot 8198 \\ &= 0.01346 \cdot 8198 \\ &\approx 110. \end{aligned}$$

Alles in allem bedeutet dies, dass die Präzision der historischen Daten durch den prädiktiven Ansatz auf $\omega_{ped} = 1.35\%$ herabgewichtet wird. Dies entspricht 110 virtuellen Kinderpatienten.

Ergebnisse der meta-analytisch prädiktiven Methode. Im Rahmen einer Sensitivitätsanalyse ist es sinnvoll, verschiedene hyper a-priori Verteilungen zu betrachten [vgl. 67, Kap. 5.7]. Die geeignete Auswahl ist von der jeweiligen Studiensituation, insbesondere von der Verteilung der zu schätzenden Parameter, abhängig. Wir betrachten für dieses Beispiel die in Abschnitt 2.5.2 beschriebenen Verteilungen. Im Idealfall hängt die a-posteriori Schätzung nicht von der gewählten hyper a-priori Verteilung ab. Die a-posteriori prädiktiven Parameter als auch N_{eff} sollten daher für jede Verteilung fast gleich sein. In den Abbildungen 4.2 und 4.3 sind die konvergenzdiagnostischen Plots für τ und θ_{ped}^* für die angewandten hyper a-priori Verteilungen dargestellt. Betrachten wir zunächst Abbildung 4.2. Anhand der Dichten lässt sich erkennen, dass sich die zwei Markovketten innerhalb einer jeden Verteilung nur wenig unterscheiden. Die Trace-Plots in 4.2a, 4.2b und 4.2c zeigen ein gutes Mixing der Markovketten. Es besteht eine hohe Autokorrelation zwischen den generierten Werten, die jedoch mit zuneh-

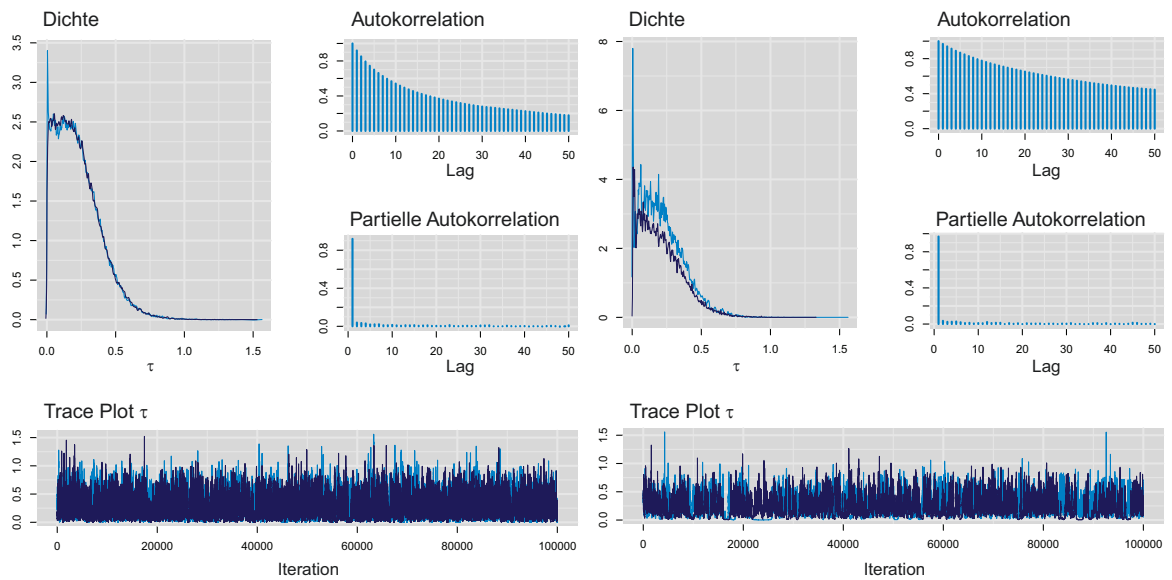
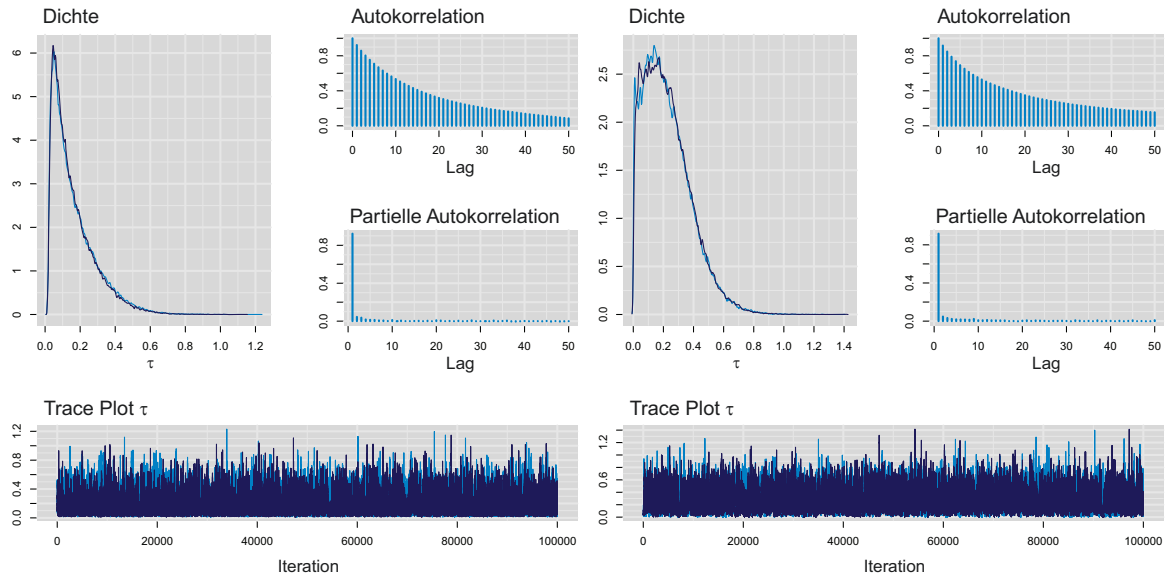


Abb. 4.2: MCMC-Diagnostiken für τ im Full-Modell der Heparin-Studien.

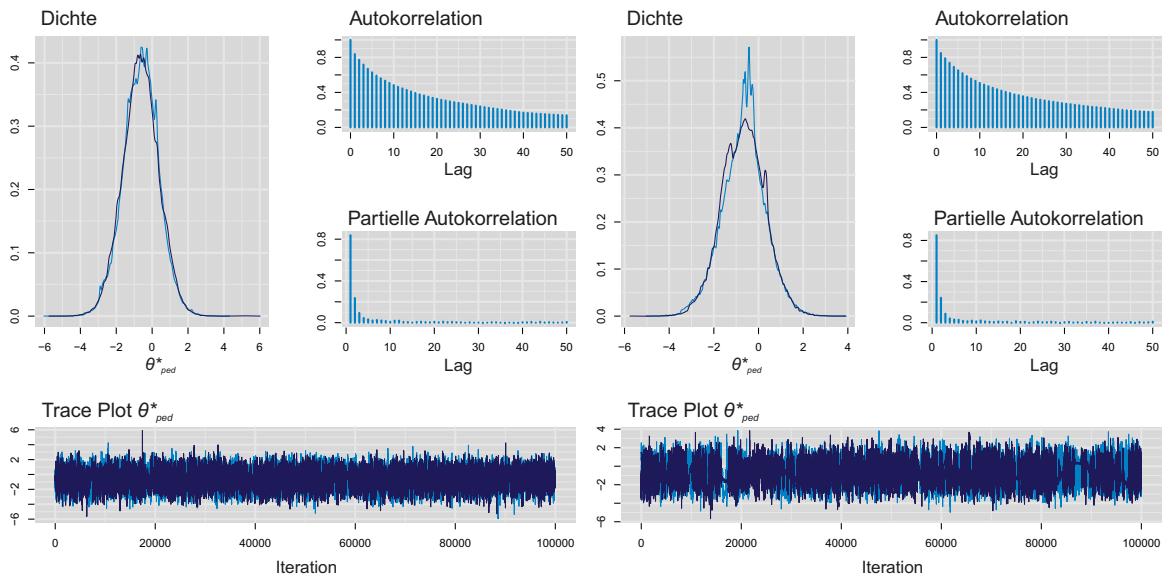
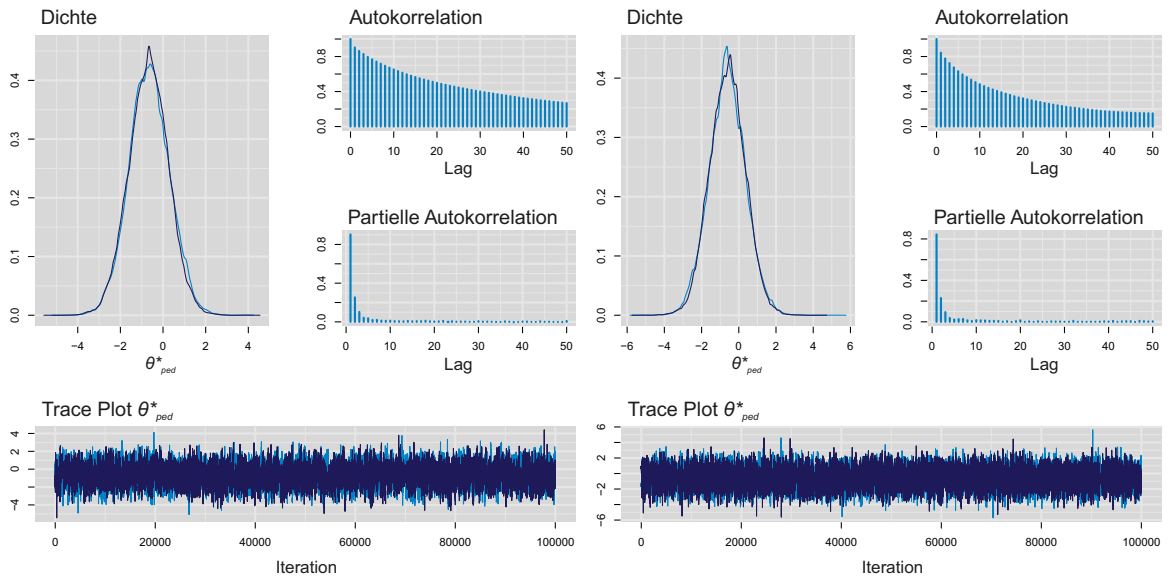


Abb. 4.3: MCMC-Diagnostiken für θ_{ped}^* im Full-Modell der Heparin-Studien.

mendem Abstand zwischen den Iterationsschritten monoton fällt. Die Half-Cauchy-Verteilung zeigt deutliche Unterschiede zu den anderen Verteilungen (vgl. Abb. 4.2d). Dies liegt zum einen daran, dass die Markovkettenstartwerte von τ bei dieser Verteilung nicht fixiert worden sind, und zum anderen die Autokorrelation extrem hoch ist. Dadurch verändern sich die Werte der Markovketten zwischen den nachfolgenden Iterationsschritten nur langsam, und es dauert länger bis die stationäre Verteilung erreicht wird. Die Inverse-Gamma-Verteilung führt im Vergleich zu den anderen Verteilungen zu einer niedrigeren Schätzung der Interstudienvariation. Wir sehen einen spitzen Gipfel der a-posteriori Dichtefunktion nahe bei 0. Dies ist das Resultat der bereits in Abschnitt 2.5.1 diskutierten Problematik, die bei der Anwendung der Inverse-Gamma-Verteilung entstehen kann.

Die Gestalt der a-posteriori prädiktiven Verteilung für θ_{ped}^* ist in allen Fällen ähnlich (vgl. Abb. 4.3). Für jede der angewandten a-priori Verteilungen ist die a-posteriori prädiktive Verteilung von θ_{ped}^* approximativ normalverteilt. Die a-posteriori Erwartungswerte sind fast identisch, $\mathbb{E}(\theta_{ped}^*|y_1 \dots y_{A+1}) = -0.615$ bis -0.667 . Die Varianzen unterscheiden sich nur geringfügig, was zu vergleichbaren Gewichtungsfaktoren und Patientenzahlen N_{eff} führt (vgl. Tab. 4.1). Aufgrund der großen historischen Fallzahl von $N_{hist} = 8198$ führen bereits kleine Abweichungen in der a-posteriori prädiktiven Varianz $V_{ped,\tau,\mu_{shift}} = \mathbb{V}\text{ar}(\theta_{ped}^*|y_1 \dots y_{A+1})$ im vollen Modell zu Unterschieden von N_{eff} .

Um die Robustheit der ermittelten Ergebnisse zu überprüfen, wenden wir die α -Trimming-Methode mit $\alpha = 0.001$ an (Kap. 3.6). Nach dem MCMC-Sampling des vollen Modells werden von den insgesamt 200 000 auswertbaren Iterationen der Markovketten diejenigen 400 Iterationen entfernt, für die die extremsten Werte für θ_{ped}^* generiert wurden. Konsequenterweise nimmt die Variation von θ_{ped}^* ab. Also ist $V_{ped,\tau,\mu_{shift}}$ im getrimmten Modell kleiner als im vollen Modell.

Aus Tabelle 4.1 wird ersichtlich, dass die virtuelle Fallzahl $N_{ped,eff}$ für eine neue Kinderstudie im Vergleich zur historischen Fallzahl $N_{hist} = 8198$ relativ niedrig ist. Es tritt ein großer Wertverlust ein. Dies ist vor allem darin begründet, dass wegen der Präsenz von nur einer Kinderstudie der Parameter θ_{ped}^* sehr unsicher geschätzt werden kann. Der systematische Verschiebungsparameter μ_{shift} zwischen den beiden Kollektiven hat dementsprechend eine große Varianz. Diese spiegelt sich wiederum direkt in der Schätzung von θ_{ped}^* wieder. Da es nur minimale Unterschiede zwischen den Werten des getrimmten und des vollen Modells gibt, sind die Schätzungen insgesamt stabil. Die Anzahl $N_{ped,eff}$ im vollen Modell variiert zwischen 105 und 117. Diese Schwankung ist im Rahmen einer Sensitivitätsanalyse relativ gering. Die historische Information wird auf ca. 1.3% herabgewichtet. Da es nur eine frühere

Heparin-Studien	A-priori Verteilung für τ							
	$\Gamma^{-\frac{1}{2}}(0.001, 0.001)$		$HN(0, 1)$		$U(0, 100)$		$HC(0, 1)$	
	α -trim		α -trim		α -trim		α -trim	
N_{hist}	8198		8198		8198		8198	
$\mathbb{E}(\mu_{shift} y)$	-0.2878	-0.2878	-0.2289	-0.2289	-0.2467	-0.2466	-0.2840	-0.2841
$\mathbb{E}(\tau y)$	0.1609	0.1606	0.2267	0.2262	0.2274	0.2269	0.2118	0.2114
Geplante Kinderstudie								
$\mathbb{E}(\theta_{ped}^* y)$	-0.663	-0.663	-0.615	-0.615	-0.634	-0.639	-0.667	-0.667
$V_{ped,\tau,\mu_{shift}}^{1/2}$	0.937	0.926	0.966	0.954	0.989	0.978	0.957	0.946
$V_{ped,\tau=0,\mu_{shift_0}}^{1/2}$	0.1120		0.1120		0.1120		0.1120	
ω_{ped}	0.0143	0.0146	0.0135	0.0138	0.0128	0.0131	0.0137	0.0140
$N_{ped,eff}$	117	120	110	113	105	107	112	115
Geplante Erwachsenenstudie								
$\mathbb{E}(\theta_{ad}^* y)$	-0.375	-0.375	-0.386	-0.386	-0.386	-0.386	-0.384	-0.384
$V_{ad,\tau,\mu_{shift}}^{1/2}$	0.2411	0.2407	0.3070	0.3063	0.3118	0.3111	0.2937	0.2931
$V_{ad,\tau=0,\mu_{shift_0}}^{1/2}$	0.1120		0.1120		0.1120		0.1120	
ω_{ad}	0.2158	0.2165	0.1331	0.1337	0.1290	0.1296	0.1454	0.1460
$N_{ad,eff}$	1769	1774	1091	1096	1057	1062	1192	1197

Tab. 4.1: Heparin-Studien: Ergebnisse der meta-analytisch prädiktiven Methode für unterschiedliche a-priori Verteilungen für τ .

A-posteriori Mittelwerte und Standardabweichungen der wichtigsten Parameter, historische Fallzahl N_{hist} , Precision Weight Effective Sample Sizes $N_{ped,eff}$ und $N_{ad,eff}$, Gewichtungsfaktoren ω_{ped} und ω_{ad} (MCMC-Simulationen: Chains=2, Iterationen=110 000, Burn-In=10 000. Getrimmtes Modell: $\alpha = 0.001$, gültige Werte=2 · 99800).

Kinderstudie gibt, ist es auch zufriedenstellend, dass die historischen Ergebnisse, die fast nur von Erwachsenenstudien stammen, wenig Gewicht bekommen. Wir können darum den systematischen Shiftparameter μ_{shift} zwischen den Kindern und Erwachsenen nur unpräzise bestimmen. Damit ist der vorhersagbare Therapieeffekt bei Kindern trotz der zusätzlichen Extrapolation der Erwachsenenendaten noch sehr unsicher.

Die a-posteriori Schätzungen mit der a-priori Inverse-Gamma-Verteilung weichen geringfügig von denen der anderen Verteilungen ab. Die Interstudienstandardabweichung $\tau = 0.1609$ ist niedriger als bei den anderen Verteilungen. Dies liegt an den von *Gelman* [21] beschriebenen Problemen, die beim MCMC-Sampling mit der Inverse-Gamma-Verteilung entstehen, wenn Werte nahe Null generiert werden. Die jeweilige Markovkette bleibt in diesem Bereich stecken, sodass die a-posteriori Verteilung sehr viel Wahrscheinlichkeitsmasse bei Null besitzt.

Konsequenterweise wird die Varianz unterschätzt. Deutlich sichtbar wird dies, wenn wir den Gewichtungsfaktor ω_{ad} für eine zukünftige Erwachsenenstudie betrachten. Während die Präzision für die Half-Normal-, Half-Cauchy- und Gleichverteilung durch ω_{ad} auf 12.90%, 13.31% bzw. 14.54% herabgewichtet wird, ist der Wertverlust bei der Inverse-Gamma-Verteilung mit 21.58% deutlich geringer. Diese unterschiedliche Gewichtung wird noch klarer bei Betrachtung der Precision Weight Effective Sample Size $N_{ad,eff}$. Während $N_{ad,eff}$ für die Half-Normal-, Half-Cauchy- und Gleichverteilung bei 1091, 1192 bzw. 1057 liegt, weicht sie bei der Inverse-Gamma-Verteilung mit 1769 stark von diesen Ergebnissen ab. Da es 18 homogene Erwachsenenstudien gibt, wird der Therapieeffekt bei Erwachsenen viel präziser vorhergesagt als bei Kindern. Daraus ergeben sich die viel höheren virtuellen Fallzahlen. Der Wertverlust im Vergleich zur Extrapolation auf eine neue pädiatrische Studie ist eindeutig geringer. Die Durchführung einer neuen Erwachsenenstudie zur Untersuchung der Wirksamkeit einer festen Dosis fraktionierten Heparins gegenüber einer adjustierten Dosis unfraktionierten Heparins scheint aufgrund der großen virtuellen Fallzahl $N_{ad,eff}$ nicht mehr nötig. Allerdings ist der historische mittlere Behandlungseffekt mit θ_{ad}^* zwischen -0.386 und -0.375 relativ gering.

Validierung. Als nächsten Schritt überprüfen wir, ob die verschiedenen Precision Effective Sample Sizes für das Kinderkollektiv tatsächlich den einzusparenden Anzahlen neuer Patienten gleichkommen. Hierzu planen wir eine neue Kinderstudie und simulieren die Therapieergebnisse (vgl. Kap. 3.3.3). In der zukünftigen Studie soll die Überlegenheit der neuen Therapie in der einseitigen Fragestellung (3.17) gezeigt werden. Als Planungsgrundlage des Therapieeffekts verwenden wir die Ereigniswahrscheinlichkeiten der beobachteten Kinderstudie die $p_T = 2/36 = 5.6\%$ und $p_C = 4/40 = 10\%$. Wir passen verschiedene Normal-Normal-Modelle unter Verwendung nicht-informativer a-priori Verteilungen (vgl. Modell (3.18)) und unter Verwendung informativer a-priori Verteilungen (vgl. Modell (3.24)) an. Die Ergebnisse sind in Tabelle 4.2 dargestellt. Um für den Nachweis des Therapieeffekts $LogOR = -0.636$ ($OR = 0.530$) eine Power von 80% zu erreichen, benötigen wir $N_{ref} = 2 \cdot 565$ auswertbare Patienten. Für die informativen Hyperparameter in der neuen Studie verwenden wir die a-posteriori prädiktiven Erwartungswerte $\mu_0 \equiv \mathbb{E}(\theta_{ped}^* | y_1, \dots, y_{A+H})$ und Varianzen $\tau_0^2 \equiv V_{ped, \tau, \mu_{shift}}$ der verschiedenen Hyperverteilungen aus dem vollen Modell. Anschließend simulieren wir 1 000 000 mal das Ergebnis der neuen Studie mit reduzierter Fallzahl $N_{ref} - N_{ped,eff}$. Ziel hierbei ist es zu kontrollieren, ob der Informationsgehalt der a-posteriori prädiktiven Verteilung der zugehörigen Precision Weight Effective Sample Size entspricht. Bei unseren Planungsmodalitäten ist dies dann der Fall, falls die bayesianische Power trotz reduzierter Fallzahl ca. 80% beträgt. Für

diesen Nachweis können wir auch die „verfälschten“ Ergebnisse des getrimmten Modells nutzen und in die Analyse mit einbringen. Der Wert der getrimmten Fallzahl und der getrimmten informativen a-priori Verteilung sollte sich ebenfalls entsprechen.

In Tabelle 4.2 sehen wir, dass die Power im schlimmsten Fall bei der Half-Normal-Verteilung auf 78.66% absinkt. Dies liegt zum einen daran, dass der a-posteriori prädiktive Erwartungswert bei -0.615 liegt. Der Behandlungseffekt in den historischen Studien wurde damit weniger stark geschätzt als der geplante Therapieeffekt -0.636 . Dementsprechend ist ein Powerverlust vorhersehbar. Zum anderen stimmt $N_{ped,eff}$ nicht mit N_{prior} überein, wie wir im nächsten Abschnitt sehen werden. Die Anzahl $N_{ped,eff}$ ist etwas zu hoch, sodass die Fallzahl zu stark reduziert wird. Die Power kann darum nicht ganz eingehalten werden. Die Unsicherheit bzw. die Varianz über den wahren zukünftigen studienspezifischen Effekt ist mit 0.989 bei der Gleichverteilung am höchsten und somit die virtuelle Fallzahl am kleinsten. Trotzdem ist der Verlust von 1.05% an Power niedrig. In allen anderen Modellen sehen wir ein Absinken der Power von ca. 1% . Absolut gesehen ist der Rückgang von 1% sehr gering und für die praktische Umsetzung vernachlässigbar. Wir können daher schließen, dass die Power in etwa eingehalten wird, obwohl die Fallzahl in der neuen Studie um die Precision Weight Effective Sample Size reduziert wird. Die Hypothese, dass die Precision Weight Effective Sample das Vorwissen, das über die informative a-priori Verteilung in das Modell der neuen Studie einfließt, repräsentiert, scheint sich vorerst zu bestätigen. An dieser Stelle sei noch einmal erwähnt, dass wir als Planungsgrundlage für den Therapieeffekt genau die beobachteten Ereignisraten der historischen Kinderstudie verwendet haben und dies somit bei der Interpretation berücksichtigen müssen. Ist der geplante Behandlungserfolg größer als der in den beobachteten historischen Daten, so kann das Powerniveau unter Verwendung von $N_{ped,prior}$ nicht eingehalten werden. Falls wir die neue Studie mit dem Normal-Normal-Modell mit nicht-informativer a-priori Verteilung durchführen und 120 Patienten weniger rekrutieren, beträgt die erzielte Power immer noch 75.2% . Ein Powerverlust ohne Verwendung informativer Prior beträgt infolgedessen maximal ca. 5% . Dies liegt an der hohen Fallzahl, die wir in diesem Studiendesign für die geplante Power von 80% benötigen.

Zunächst können wir an dieser Stelle zusammenfassend notieren, dass es so scheint, als ob die Precision Weight Effective Sample Size auch tatsächlich den Wert der informativen a-priori Parameter wiedergibt. Ob dies wirklich so ist, können wir aufgrund der großen benötigten Fallzahl aber noch nicht eindeutig sagen. Die Antwort geben wir im nachfolgenden Kapitel.

Heparin-Studien		Normal-Normal-Modell mit nicht-informativer und informativer a-priori Verteilung			
		Hyper a-priori Verteilung für τ bzw. τ^2 im vollen Modell			
		$\tau^2 \sim \Gamma(0.001, 0.001)$	$\tau \sim HN(0, 1)$	$\tau \sim U(0, 100)$	$\tau \sim HC(0, 1)$
		α -trim	α -trim	α -trim	α -trim
Angenommene		interventionelle Therapie		$p_T = 2/36 = 5.6\%$	
Ereignisraten		konventionelle Therapie		$p_C = 4/40 = 10\%$	
A-priori Verteilung	μ_0	0	0	-0.663	-0.663
$\theta \mu_0, \tau_0^2 \sim N(\mu_0, \tau_0^2)$	τ_0	∞	∞	0.937	0.926
Fallzahl	2-565 =1130	2-505 =1010	1130 =1130	1130 =1130	1130 =1130
Bayesianische Power	80.0%	75.2%	79.10%	78.99%	78.74%
			78.66%	78.95%	79.01%
				79.20%	79.19%
				0.957	0.946
				-0.667	-0.667

Tab. 4.2: Heparin-Studien: Poweranalyse zur Validierung der errechneten Precision Weight Effective Sample Size.

Normal-Normal-Modell mit nicht-informativen und informativen a-priori Verteilungen für den Studieneffekt der neuen Kinderstudie. Die informativen Parameter sind der a-posteriori prädiktive Erwartungswert $\mathbb{E}(\theta_{ped}^* | y)$ und die a-posteriori prädiktive Varianz $V_{ped, \tau, \mu_{shif t}}$ aus dem jeweiligen Modell der meta-analytisch prädiktiven Methode. Bei den informativen a-priori Verteilungen ist die Fallzahl um $N_{ped, eff}$ reduziert.

Prior Effective Sample Size. Zusätzlich zur Precision Weight Effective Sample Size N_{eff} berechnen wir die Prior Effective Sample Size N_{prior} für eine neue Kinderstudie und eine neue Erwachsenenstudie. Wir können dadurch überprüfen, inwiefern sich die Precision Weight Effective Sample Size nach *Neuenschwander et al.* [50] von der Prior Effective Sample Size, die die tatsächliche Fallzahleinsparung durch Verwendung der informativen a-priori Parameter repräsentiert, unterscheidet. Wir planen die neue Kinderstudie unter den bereits erwähnten Annahmen. Allerdings simulieren wir die Power nicht, sondern berechnen sie nach den Formeln in Kapitel 3.4. Konsequenterweise ändert sich dadurch geringfügig die benötigte Referenzfallzahl ohne Nutzung des Vorwissen im Vergleich zum vorherigen Kapitel. Zum Nachweis von $p_T=5.6\%$ vs. $p_C=10\%$ in einer neuen Kinderstudie benötigen wir $N_{ped,ref} = 1172$ neue Patienten. In einer neuen Erwachsenenstudie, die wir entsprechend den gepoolten Gesamtergebnisraten $p_{TC} = (151 + 211)/(4181 + 3941) = 4.46\%$ und dem geschätzten Log Odds Ratio -0.386 planen, brauchen wir $N_{ad,ref} = 5112$ erwachsene Studienteilnehmer. In Tabelle 4.3 sind Voraussetzungen und die Ergebnisse aufgeführt. In dieser konkreten neuen Kinderstudie können wir zwischen 81 und 97 Patienten einsparen. Die Prior Effective Sample Size $N_{ped,prior}$ fällt insgesamt etwas kleiner aus als die Precision Weight Effective Sample Sizes $N_{ped,eff}$ (vgl. Tab. 4.1). Die bereits beschriebenen Unterschiede von $N_{ped,eff}$ bezüglich den im vollen Modell verwendeten hyper a-priori Verteilungen treten für $N_{ped,prior}$ im gleichen Umfang auf. Die informativen a-priori Parameter sind logischerweise identisch, sodass die Auswirkungen auf die Fallzahlen bestehen bleiben. Die Inverse-Gamma-Verteilung schätzt die a-posteriori Varianz am kleinsten. Die a-priori Parameter sind somit am informativsten und darum $N_{ped,prior}$ bzw. $N_{ad,prior}$ am größten. Das Verhältnis der einsparbaren Patienten $N_{ped,prior}$ zur historischen Fallzahl $N_{hist} = 8198$ entspricht für diese konkrete Kinderstudie ca. 1%. Da die Precision Weight Effective Sample Size $N_{ped,eff}$ geringfügig größer ist als die Prior Effective Sample Size $N_{ped,prior}$, ist auch das Absinken der Power um 1% bei Reduzierung der Fallzahl um $N_{ped,eff}$ im Nachhinein zu erklären (vgl. Tab.4.2). Es werden zu viele Patienten eingespart, sodass das Powerlevel von 80% nicht ganz eingehalten werden. Die Precision Weight Effective Sample Size und die Prior Effective Sample Size sind in diesem Beispiel nicht gleich. Ein etwas anderes Bild ergibt sich bei der neuen Erwachsenenstudie. Insgesamt ist hier $N_{ad,prior}$ um einiges höher als die im Vorfeld aus den Präzisionen errechnete Fallzahl $N_{ad,eff}$. Es können somit viel mehr Patienten eingespart werden. Die relative Ersparnis durch Nutzung des Vorwissen beträgt zwischen 26% und 41%. Die großen Unterschiede in $N_{ad,prior}$ bezüglich der hyper a-priori Verteilungen zeigen sich auch hier. Insgesamt besitzen die historischen Studienergebnisse mehr Informationsgehalt für eine neue Erwachsenenstudie als für

eine Kinderstudie. Dies ergibt sich aber unmittelbar aus der unterschiedlichen Zusammensetzung der historischen Studien mit 18 Erwachsenenstudien gegenüber einer Kinderstudie. Zwischen den getrimmten und ungetrimmten Prior Effective Sample Sizes bestehen die zu erwartenden sehr kleinen Unterschiede, die allerdings vernachlässigbar sind. Aufgrund der hohen benötigten Fallzahl für eine neue Erwachsenenstudie zum Nachweis dieses Therapieeffekts ist eine tatsächliche Durchführung der Studie mit dem angenommenen Behandlungseffekt $\text{LogOR} = -0.386$ eher unrealistisch.

Wir können festhalten, dass die Prior Effective Sample Size N_{prior} immer stark von der ohne Vorwissen benötigten Fallzahl N_{ref} und den informativen a-priori Parametern der neuen Studie abhängt. Der meta-analytisch prädiktive Ansatz funktioniert in diesem Beispiel gut. Die Unterschiede zwischen den Ergebnissen der hyper a-priori Verteilungen für eine Kinderstudie sind gering, für eine Erwachsenenstudie etwas größer. Die Inverse-Gamma-Verteilung führt zu stärker abweichenden Resultaten im Vergleich zu den anderen betrachteten Verteilungen. Insgesamt liegt die Ersparnis an Kinderpatienten durch die Extrapolation der Resultate der historischen Erwachsenenstudien und der einen Kinderstudie im Bereich zwischen 6.9% und 8.3%. Bei einer neuen Erwachsenenstudie kann trotz der hohen benötigten Gesamtfallzahl die Anzahl neuzurekrutierender Patienten um 15.7% bis 34.5% reduziert werden.

Heparinstudien	Normal-Normal-Modell mit informativer a-priori Verteilung							
	Hyper a-priori Verteilung für τ bzw. τ^2 im vollen Modell							
	$\Gamma^{-\frac{1}{2}}(0.001, 0.001)$		$HN(0, 1)$		$U(0, 100)$		$HC(0, 1)$	
	α -trim		α -trim		α -trim		α -trim	
Geplante Kinderstudie	$p_T=2/36=5.6\%$				$p_C=4/40=10\%$			
	logOR=-0.636							
$N_{ped,ref}$	1172							
$\mu_{ped,0}$	-0.663	-0.663	-0.615	-0.615	-0.634	-0.639	-0.667	-0.667
$\tau_{ped,0}$	0.937	0.926	0.966	0.954	0.989	0.978	0.957	0.946
$N_{ped,prior}$	97	99	81	83	81	84	93	96
$N_{ped,prior}/N_{ped,ref}$	0.083	0.084	0.069	0.071	0.069	0.072	0.080	0.082
$N_{ped,prior}/N_{hist}$	0.012	0.012	0.010	0.010	0.010	0.010	0.011	0.012
Geplante Erwachsenenstudie	$p_T=3.64\%$				$p_C=5.3\%$			
	logOR=-0.386							
$N_{ad,ref}$	5112							
$\mu_{ad,0}$	-0.375	-0.375	-0.386	-0.386	-0.386	-0.386	-0.384	-0.384
$\tau_{ad,0}$	0.2411	0.2407	0.3070	0.3063	0.3118	0.3111	0.2937	0.2931
$N_{ad,prior}$	2175	2183	1376	1382	1332	1339	1495	1501
$N_{ad,prior}/N_{ad,ref}$	0.426	0.427	0.269	0.270	0.261	0.262	0.293	0.294
$N_{ad,prior}/N_{hist}$	0.265	0.266	0.168	0.169	0.162	0.163	0.182	0.183

Tab. 4.3: Heparin-Studien: Bestimmung der Prior Effective Sample Size N_{prior} für eine neue Studie unter Verwendung der Parameterschätzungen des vollen Modells für unterschiedliche a-priori Verteilungen für τ .

Informative a-priori Parameter der neuen Studie $\mu_{ped,0}$, $\mu_{ad,0}$, $\tau_{ped,0}$ und $\tau_{ad,0}$, benötigte Fallzahl mit nicht-informativen a-priori Verteilungen $N_{ped,ref}$ und $N_{ad,ref}$, Prior Effective Sample Sizes $N_{ped,prior}$ und $N_{ad,prior}$. Normal-Normal-Modell mit nicht-informativen und informativen a-priori Verteilungen für den Studieneffekt einer neuen Kinder- und Erwachsenenstudie. Die informativen Parameter der Kinderstudie sind der a-posteriori prädiktive Erwartungswert $\mu_{ped,0} = \mathbb{E}(\theta_{ped}^*|y)$ und die a-posteriori prädiktive Varianz $\tau_{ped,0}^2 = V_{ped,\tau,\mu_{shift}}$ aus dem vollen Modell der meta-analytisch prädiktiven Methode. Für die Erwachsenenstudie sind die informativen a-priori Parameter analog $\mu_{ad,0} = \mathbb{E}(\theta_{ad}^*|y)$ und $\tau_{ad,0}^2 = V_{ad,\tau,\mu_{shift}}$.

4.2 Metformin-Studien

Studienbeschreibung. Im zweiten Beispiel betrachten wir klinische Studien mit einer stetigen normalverteilten Zielgröße. Hierbei handelt es sich um Studien, die den Einfluss einer Metformin-Monotherapie bei Patienten mit Typ II Diabetes Mellitus untersuchen. Der Wirkstoff Metformin hemmt die Glucose-Abgabe der Leber und steigert den peripheren Glucose-Verbrauch [42, S. 446f.]. Die Wirksamkeit wird durch die Veränderung des Blutzuckerwertes Glykohämoglobin (HbA1c) beschrieben. Die Zielgröße ist die standardisierte Mittelwertdifferenz des HbA1c-Werts (in %) zwischen Baseline und Follow-up. Als Datengrundlage dienen die Ergebnisse eines Cochrane-Reviews von *Saenz et al.* [56] mit insgesamt 28 Studien an 1989

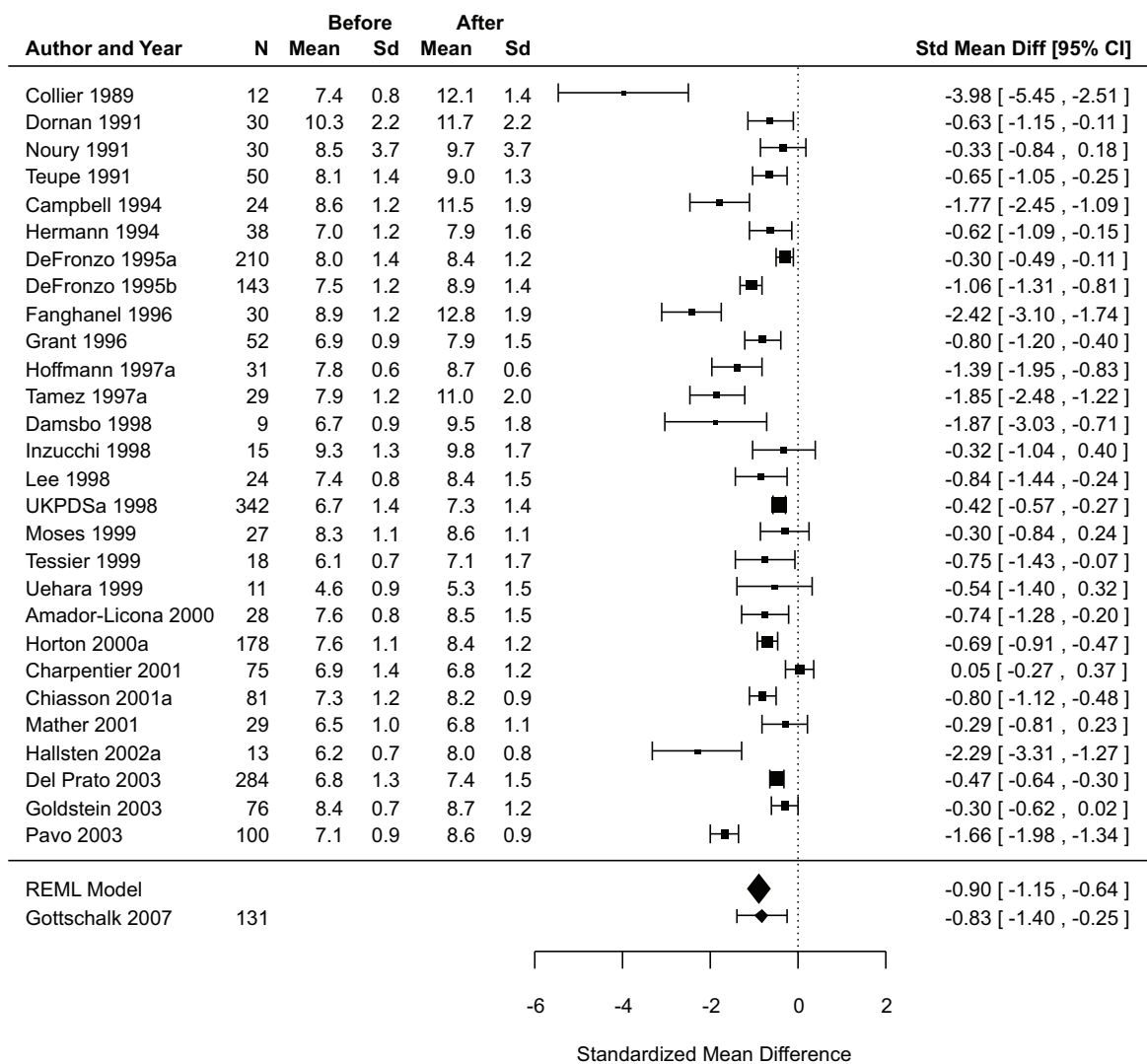


Abb. 4.4: Forestplot der Metformin-Studien.

Heterogenität (Erwachsenenstudien): $\tau^2 = 0.3876$ (95% CI 0.2551, 1.1479), $df=27$ ($p<0.001$), $I^2 = 92.26\%$ (95% CI 88.69, 97.25%). Test für Gesamteffekt (Erwachsenenstudien): $p<0.001$.

erwachsenen Patienten sowie eine einzige Studie an 131 Kindern von *Gottschalk et al.* [26]. Die Outcomes der einzelnen Studien sind in Abbildung 4.4 dargestellt. Bei den Erwachsenen ist eine gepoolte mittlere Veränderung von -0.90 [95% CI -1.15, -0.64] im HbA1c-Wert (%) zu beobachten. Die Kinderstudie zeigt mit einer mittleren Differenz von -0.83 [95% CI -1.40, -0.25] einen vergleichbaren Effekt. Die Erwachsenenstudien sind im Gegensatz zum Heparin-Beispiel 4.1 sehr heterogen ($I^2 = 92.26\%$). Man beachte, dass es sich um standardisierte Effekte handelt, also $\frac{1}{\sigma} \sum_{i=1}^n (x_{1,i} - x_{2,i})$, wobei $x_{1,i}$ der Baselinewert und $x_{2,i}$ der Follow-up-Wert des Patienten i ist. Die mittleren HbA1c-Differenzen streuen sehr stark zwischen den Studien. Die mittlere Effektstärke liegt zwischen den Extremfällen -3.98 (*Collier 1989*), der eine enorme Wirkung der Therapie suggeriert, und 0.05 (*Charpentier 2001*), der keinen Behandlungserfolg darstellt. Die Intrastudienvarianz haben wir aus den berichteten Konfidenzintervallen berechnet. Aufgrund der Standardisierung ist sie überall ähnlich und wird nur durch die unterschiedlichen Fallzahlen der einzelnen Studien beeinflusst. Die Studien überdecken eine große Zeitspanne, was die Vergleichbarkeit der einzelnen Ergebnisse beeinträchtigen kann. Auffällig ist, dass die früheren Studien bis zum Jahr 1998 einen tendenziell größeren Therapieeffekt als die späteren Studien zeigen.

Ergebnisse der meta-analytisch prädiktiven Methode. Unser Ziel ist es, die a-posteriori prädiktiven Verteilungen der mittleren HbA1c-Veränderung für eine zukünftige Studie bei Kindern θ_{ped}^* und bei Erwachsenen θ_{ad}^* zu bestimmen. Um die in Kapitel 3 beschriebenen hierarchischen Modelle anzuwenden, müssen wir in diesem Beispiel einschränkend voraussetzen, dass die beobachtete Intrastudienvarianz der wahren Intrastudienvarianz σ^2 entspricht. Die historischen Ergebnisse sind aufgrund der Heterogenität der Studien aber nur sehr eingeschränkt für eine neue Studie verwendbar. In Tabelle 4.4 sind die Ergebnisse des meta-analytisch prädiktiven Ansatzes und die Gewichtung der Präzision dargestellt. Aufgrund der sehr großen Standardabweichung zwischen den Studien von $\tau > 0.63$ für alle verwendeten hyper a-priori Verteilungen, können sowohl θ_{ped}^* als auch θ_{ad}^* nur sehr unpräzise vorhergesagt werden. Die bayesschen Schätzungen von τ stimmen ungefähr mit dem frequentistischen Ergebnis $\tau = 0.62$ für die Erwachsenenstudien überein. Die Standardabweichungen $V_{ped,\tau,\mu_{shift}}^{1/2}$ und $V_{ad,\tau,\mu_{shift}}^{1/2}$ der zukünftigen studienspezifischen Effekte, also die der vorhergesagten mittleren HbA1c-Senkungen, sind größer als 0.92 bzw. größer als 0.66. Die Schätzungen aller Parameter für alle hyper a-priori Verteilungen für τ sind fast identisch. Die Unterschiede zwischen den vollen Modellen und den getrimmten Modellen sind ebenfalls marginal. Eine Entfernung von extremen Iterationsschritten in den generierten Markovketten hat damit

Metformin-Studien	A-priori Verteilung für τ							
	$\Gamma^{-\frac{1}{2}}(0.001, 0.001)$		$HN(0, 1)$		$U(0, 100)$		$HC(0, 1)$	
	α -trim		α -trim		α -trim		α -trim	
N_{hist}	2120		2120		2120		2120	
$\mathbb{E}(\mu_{shift} y)$	0.0720	0.0720	0.0780	0.0780	0.0765	0.0764	0.0725	0.0725
$\mathbb{E}(\tau y)$	0.6373	0.6368	0.6508	0.6503	0.6623	0.6619	0.6475	0.6471
Geplante Kinderstudie								
$\mathbb{E}(\theta_{ped}^* y)$	-0.8252	-0.8252	-0.8236	-0.8237	-0.8240	-0.8240	-0.8267	-0.8266
$V_{ped, \tau, \mu_{shift}}^{1/2}$	0.9225	0.9094	0.9407	0.9274	0.9609	0.9471	0.9352	0.9220
$V_{ped, \tau \equiv 0, \mu_{shift_0}}^{1/2}$	0.031		0.031		0.031		0.031	
ω_{ped}	0.0011	0.0012	0.0011	0.0011	0.0010	0.0011	0.0011	0.0011
$N_{ped, eff}$	2	2	2	2	2	2	2	2
Geplante Erwachsenenstudie								
$\mathbb{E}(\theta_{ad}^* y)$	-0.8972	-0.8984	-0.8988	-0.8986	-0.9003	-0.9001	-0.8972	-0.8970
$V_{ad, \tau, \mu_{shift}}^{1/2}$	0.6662	0.6629	0.6763	0.6757	0.6906	0.6902	0.6732	0.6727
$V_{ad, \tau \equiv 0, \mu_{shift_0}}^{1/2}$	0.031		0.031		0.031		0.031	
ω_{ad}	0.0022	0.0022	0.0021	0.0021	0.0020	0.0021	0.0021	0.0021
$N_{ad, eff}$	4	4	4	4	4	4	4	4

Tab. 4.4: Metformin-Studien: Ergebnisse der meta-analytisch prädiktiven Methode für unterschiedliche a-priori Verteilungen für τ .

A-posteriori Mittelwerte und Standardabweichungen der wichtigsten Parameter, historische Fallzahl N_{hist} , Precision Weight Effective Sample Sizes $N_{ped, eff}$ und $N_{ad, eff}$, Gewichtungsfaktoren ω_{ped} und ω_{ad} (MCMC-Simulationen: Chains=2, Iterationen=110 000, Burn-In=10 000. Getrimmtes Modell: $\alpha = 0.001$, gültige Werte=2 · 99800).

keine essentielle Auswirkung auf die Parameterschätzungen. Zudem ist der Einfluss der gewählten a-priori Verteilungen in den a-posteriori Verteilungen im Gegensatz zum Gewicht der beobachteten Daten in Form der Likelihood sehr gering. Wir können daraus schließen, dass die Schätzungen der Parameter fast identisch und damit robust sind. Demnach ist es in diesem Beispiel egal, welche hyper a-priori Verteilung wir anwenden. Für die eventuelle Planung einer neuen Studie würden wir dann die am wenigsten informative Verteilung, in diesem Fall die Gleichverteilung, benutzen.

Betrachten wir die ermittelten Precision Weight Effective Sample Sizes, so ist das Resultat allerdings ernüchternd (Tab. 4.4). Die Präzision des Behandlungseffekts einer neuen Kinderstudie wird aufgrund der Heterogenität auf $\omega_{ped} = 0.1\%$ herabgewichtet. In Form der Fallzahl entspricht diese Gewichtung einer Reduzierung von insgesamt $N_{hist} = 2120$ Patienten auf

entsprechend $N_{ped,eff} = 2$ Patienten. Die historischen Ergebnisse bringen damit keinen Erkenntnisgewinn über einen zukünftig zu erwartenden Therapieeffekt. Positiv ist jedoch, dass das Resultat bei allen verwendeten Hyperverteilungen gleich ist.

Für die Verwendung für eine neue Studie an Erwachsenen verdoppelt sich der Wert der historischen Daten auf ein Präzisionsgewicht von $\omega_{ad} = 0.2\%$ bzw. $N_{ad,eff} = 4$. Dass die Güte der Daten insgesamt so schlecht bewertet wird, liegt an der sehr hohen Heterogenität und hohen Interstudienvariation. Zusammenfassend können wir festhalten, dass eine Extrapolation der historischen Information in diesem Anwendungsbeispiel nicht hilfreich ist. Es ist sinnvoller in der neuen Studie auf das Vorwissen zu verzichten. Wie es bei frequentistisch geplanten Studien üblich ist, können wir den mittleren gepoolten Effekt ausschließlich als Planungsgrundlage für den tendenziell zu erwarteten Therapieeffekt in einer komplett neuen Studie verwenden. Der historische Effekt ist somit nur ein Indiz für den zu erwartenden Behandlungseffekt in der zukünftigen Studie.

Auf eine Validierung der Precision Weight Effective Sample Size und eine Poweranalyse verzichten wir in diesem Beispiel. Die geplante Power wird sehr wahrscheinlich eingehalten, da eine Veränderung der geplanten Fallzahl um 2 Patienten keine große Wirkung auf die Power haben kann. Genauso macht es keinen Sinn die Prior Effective Sample Size für eine spezielle neue Studie zu bestimmen, da die informativen a-priori Parameter in jedem Fall zu uninformativ sind.

Aus medizinischer Perspektive könnten wir zu der Ansicht gelangen, dass die bisherigen Studien aufgrund des großen Zeitraums (1989 – 2007) nicht direkt miteinander vergleichbar sind. Dies könnte beispielsweise an Therapieoptimierungen oder an Gesetzesänderungen zur Durchführung von klinischen Studien liegen (vgl. Kap. 1.2). Die Annahme der Austauschbarkeit der Studien könnte damit verletzt sein. Darum berücksichtigen wir exemplarisch einmal nur die Studien ab 1999 als historische Datengrundlage. Die Gesamtfallzahl reduziert sich auf $N_{hist}=1051$. Unter Anwendung der meta-analytisch prädiktiven Methode erhalten wir wie zuvor $\omega_{ped} = 0.1\%$ bzw. $\omega_{ad} = 0.2\%$ und $N_{ped,eff} = 2$ bzw. $N_{ad,eff} = 4$ virtuelle Patienten. Das bedeutet, dass trotz Halbierung der historischen Fallzahl N_{hist} , die umgerechnete Anzahl virtueller Patienten bzw. die Gewichtungsfaktoren nicht weiter gesunken ist. Der Informationsgehalt dieser Studien ab 1999 ist doppelt so hoch wie der aller Studien zusammen. Allerdings sind die absoluten Gewichtungsfaktoren und die dementsprechenden Anzahlen virtueller Patienten immer noch sehr enttäuschend niedrig. Für weitere reelle Planungen sind die historischen Metformin-Studien also nicht brauchbar.

4.3 FeNO-Studien

Studienbeschreibung. Als letztes Beispiel betrachten wir klinische Studien an Patienten mit Asthma bronchiale. In den interventionellen Behandlungsgruppen wird die Medikation basierend auf Stickstoffmonoxid-Messungen eingestellt. FeNO ist die Abkürzung für fraktioniertes exhalierendes Stickstoffmonoxid. Bei der FeNO-Messung wird die Konzentration an Stickstoffmonoxid in der ausgeatmeten Luft gemessen. Stickstoffmonoxid wird in den Atemwegen und dort unter anderem von Entzündungszellen gebildet, die einen wesentlichen Einfluss an der Entstehung und Andauer von Asthma bronchiale haben. Mittels eines FeNO-Monitorings kann das Entzündungsausmaß bestimmt und daraufhin die Steroiddosis angepasst werden [6].

Als Datenbasis verwenden wir die Publikation „*A systematic review and meta-analysis: tailoring asthma treatment on eosinophilic markers (exhaled nitric oxide or sputum eosinophils)*“ von *Petsky et al.* [52]. Ein Ziel dieses Reviews ist es, die Effektivität der Einstellung der Behandlungsstrategie basierend auf dem ausgeatmeten Stickstoffmonoxid zu beurteilen. Als Kontrollgruppe dienen pädiatrische und erwachsene Patienten mit klinischen Symptomen mit oder ohne Spirometrie bzw. Peak-Flow-Messung. Der Endpunkt ist auf eine binäre Zielgröße, die Verschlechterung in den asthma-assoziierten Erfolgsparametern im Studienverlauf, redu-

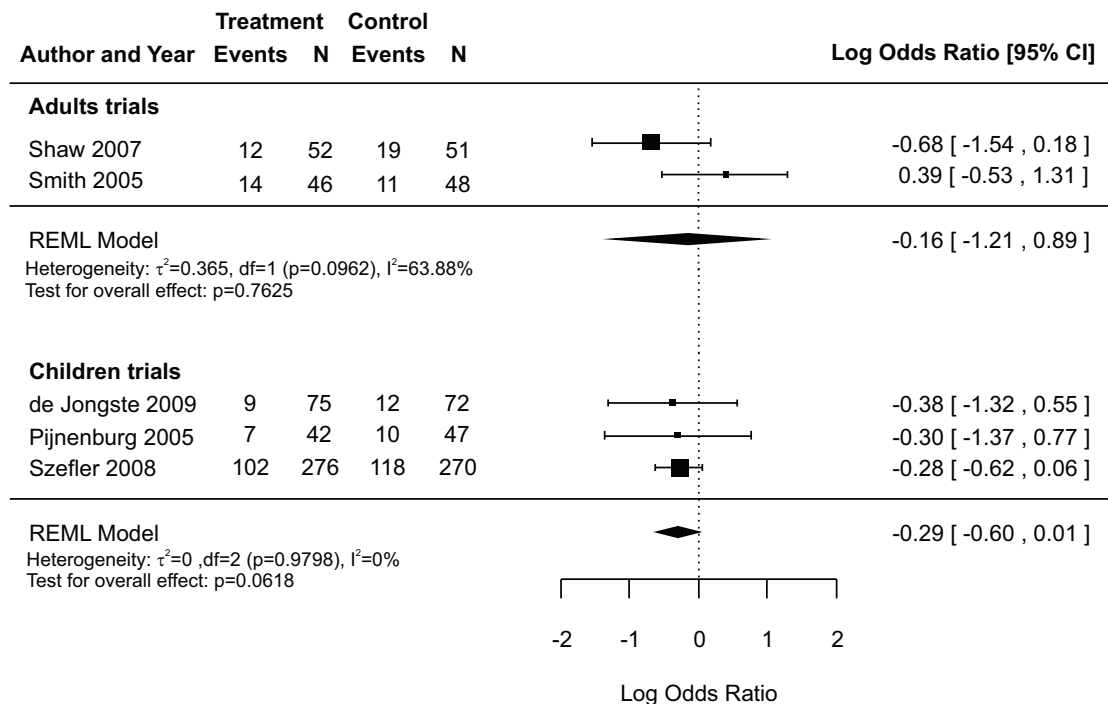


Abb. 4.5: Forestplot der FeNO-Studien.

ziert. Die einzelnen Asthmaparameter sind im Detail in *Petsky et al.* [52, S. 200f.] aufgelistet. In die Auswertung sind die Ergebnisse von 2 Studien mit insgesamt 197 Erwachsenen und von 3 pädiatrischen Studien mit 782 Kindern eingeschlossen. In Abbildung 4.6 sind die beobachteten Outcomes der einzelnen Studien dargestellt. Die gepoolten Effekte in den einzelnen Kollektiven berechnen wir mittels eines frequentistischen Random-Effects-Modells (Restricted Maximum-Likelihood (REML)). Als gepoolten Effekt erhalten wir bei den Erwachsenen ein Odds Ratio von 0.85 [95% CI 0.30, 2.43], bei den Kindern ein Odds Ratio von 0.75 [95% CI 0.55, 1.01]. Die beiden Erwachsenenstudien zeigen sehr heterogene Ergebnisse ($I^2 = 63.88\%$) mit sogar gegenläufigen Effekten. Entsprechend weit ist auch das 95%-Konfidenzintervall des Effektschätzers. Im Gegensatz dazu sind die mittleren Effekte der Kinderstudien fast identisch und die Studien in sich sehr homogen ($I^2 = 0\%$). Der gepoolte Effekt ist knapp nicht signifikant. Selbstverständlich stellt dieses Beispiel nicht die geläufige Situation klinischer Studien dar. In der Regel gibt es viele Studien an Erwachsenen und nur wenige an Kindern. Wenn wir in der Rolle des Studienplaners solche gegensätzlichen Studienresultate (Abb. 4.5) wie bei den Erwachsenen beobachten, würden wir uns auf die Extrapolation der Kinderdaten beschränken. Die Beobachtungen der Erwachsenen blieben für eine neue Studie an Kindern unberücksichtigt. Wir werden hier jedoch aus theoretischen Gesichtspunkten exemplarisch die kombinierten Daten beider Studienpopulationen nutzen. Wenn wir einen austauschbaren Zusammenhang zwischen den Studieneffekten voraussetzen (vgl. Kap. 2.3.1(ii)), könnten die beiden Resultate an Erwachsenen eventuell mögliche Realisierungen im Rahmen der Modellannahme repräsentieren. Das FeNO-Beispiel soll dazu dienen, die Grenzen des meta-analytisch prädiktiven Ansatzes mit zwei Kollektiven beim Vorliegen weniger unsicherer historischer Daten aufzuzeigen. Des Weiteren ist das Beispiel die Motivation für die genauere Untersuchung dieser Situation in simulierten Studienszenarien (Kap. 5).

Ergebnisse der meta-analytisch prädiktiven Methode. Als ersten Schritt simulieren wir wieder das volle Modell (3.2), um die a-posteriori prädiktive Verteilung des studienspezifischen Log Odds Ratio zu bestimmen. Hierfür verwenden wir die MCMC-Einstellungen aus Kapitel 3.5. In Abbildung 4.6 sind die Dichten der verschiedenen hyper a-priori Verteilungen und die resultierenden a-posteriori Dichten für τ dargestellt. Auf den ersten Blick lassen sich deutliche Unterschiede erkennen. Die schwach-informativen Half-Normal- und Half-Cauchy-Hyperverteilungen liefern fast identische a-posteriori Ergebnisse. Die noch weniger informativ Gleichverteilung verhält sich etwas konservativer und legt mehr Wahrscheinlichkeitsmasse auf extremere Werte für τ . Bei diesen drei Hyperverteilungen lassen sich zwischen dem unge-

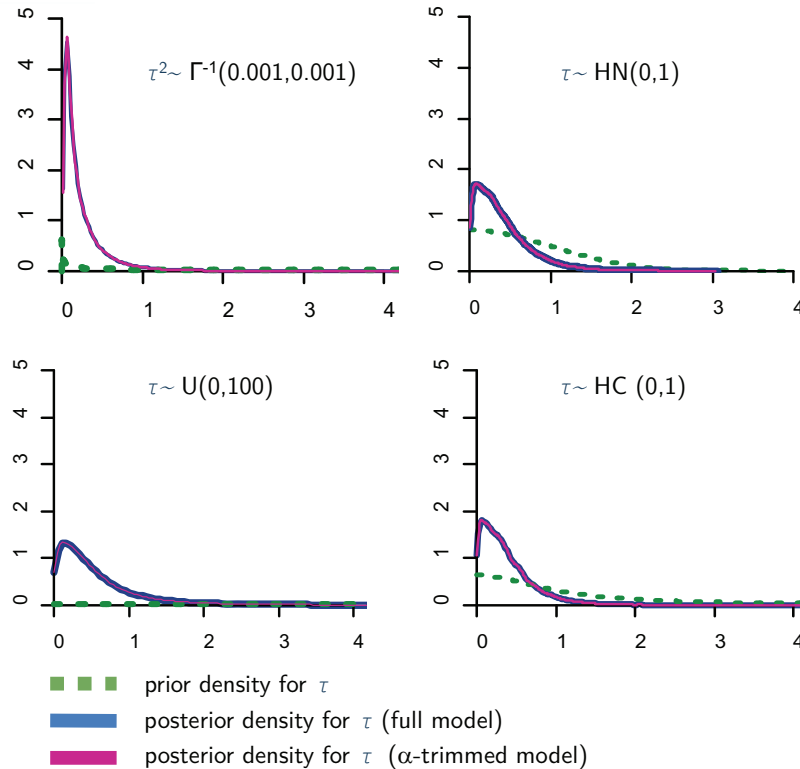
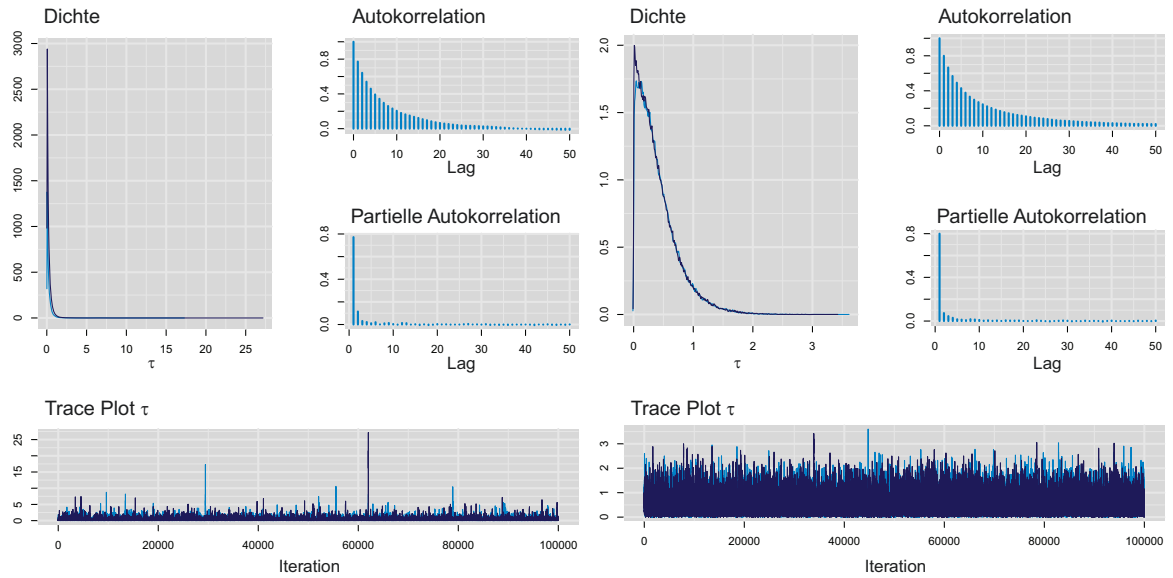
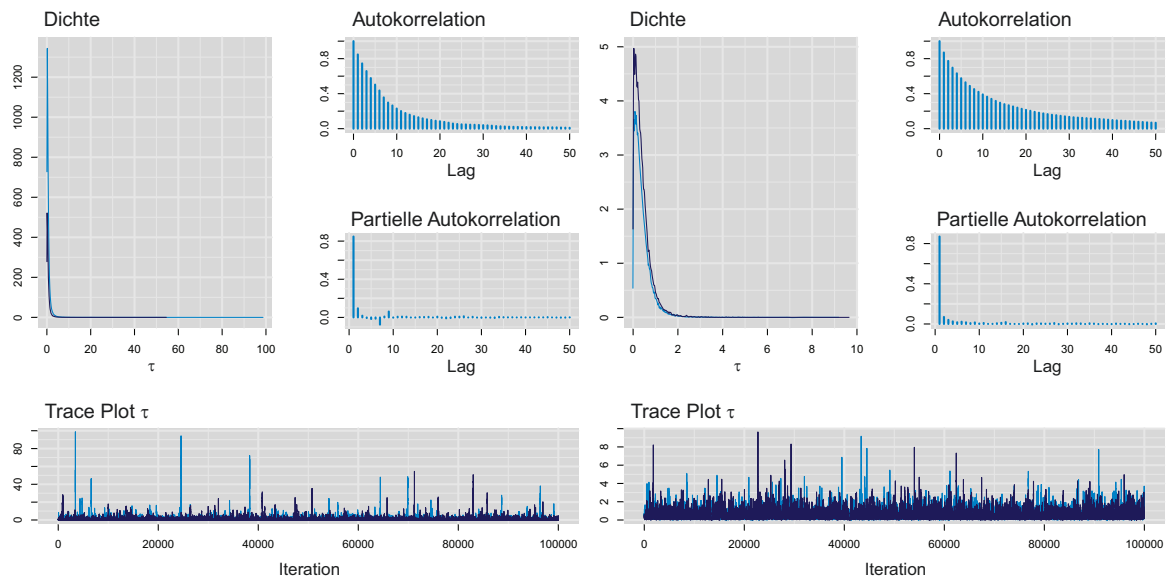


Abb. 4.6: A-priori und a-posteriori Dichtefunktionen der Interstudienstandardabweichung τ für die FeNO-Studien.

trimmten und getrimmten Modell keine Unterschiede für τ erkennen. Anders verhält sich die Schätzung von τ bei der Inverse-Gamma-Verteilung. Hier erkennen wir, dass deutlich mehr Wahrscheinlichkeitsmasse nahe bei 0 liegt. Die getrimmte a-posteriori Verteilung verläuft etwas spitzer als die ungetrimmte. Die a-priori Inverse-Gamma-Verteilung (grün) ähnelt der Gleichverteilung, mit der Ausnahme eines kleinen Spikes bei 0. Durch die wenigen historischen Beobachtungen bewirkt dieser Spike, dass die a-posteriori Verteilung kleineren Werten mehr Wahrscheinlichkeit zuordnet. In den diagnostischen Plots 4.7 des vollen Modells bestätigt sich dies. Man beachte die unterschiedlichen Skalierungen in den einzelnen Abbildungen. Die Autokorrelation ist bei allen Verteilungen ähnlich und nimmt mit fortschreitenden Iterationen schnell monoton ab. Die einzelnen Iterationsschritte der Markovketten lassen sich am anschaulichsten in den Traceplots visualisieren. Das beste Mixing der zwei Ketten sehen wir bei der Half-Normal-Verteilung. Die Iterationenswerte für τ überschreiten fast nie den Wert 3. Bei der Half-Cauchy-Verteilung verhält es sich ähnlich, auch wenn die einzeln auftretenden Maxima bereits deutlich größer sind. Viel unbeständiger sieht dies bei der unbeschränkten Inverse-Gamma-Verteilung aus. Wir sehen, dass immer wieder einzelne Ausreißer, beispielsweise $\tau > 10$, generiert werden. Allerdings beinhaltet der Großteil der Iterationen sehr kleine Werte, was trotz der vereinzelt Extremwerte zu einer insgesamt sehr kleinen mittleren

(a) A-priori Verteilung: $\tau^2 \sim \Gamma^{-1}(0.001, 0.001)$.(b) A-priori Verteilung: $\tau \sim HN(0, 1)$.(c) A-priori Verteilung: $\tau \sim U(0, 100)$.(d) A-priori Verteilung: $\tau \sim HC(0, 1)$.**Abb. 4.7:** MCMC-Diagnostiken für τ im Full-Modell der FeNO-Studien.

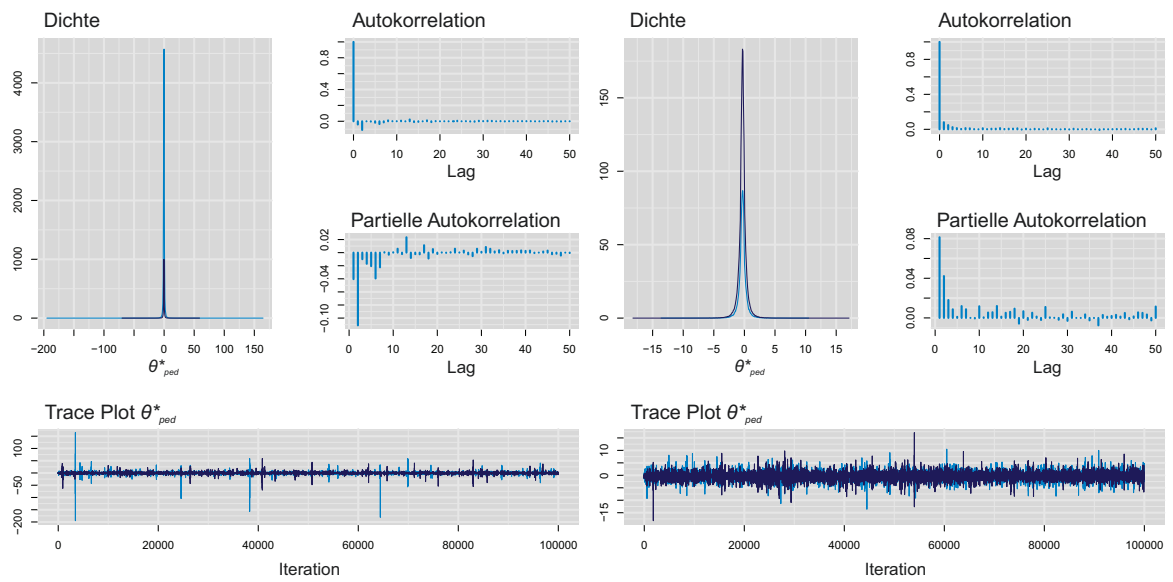
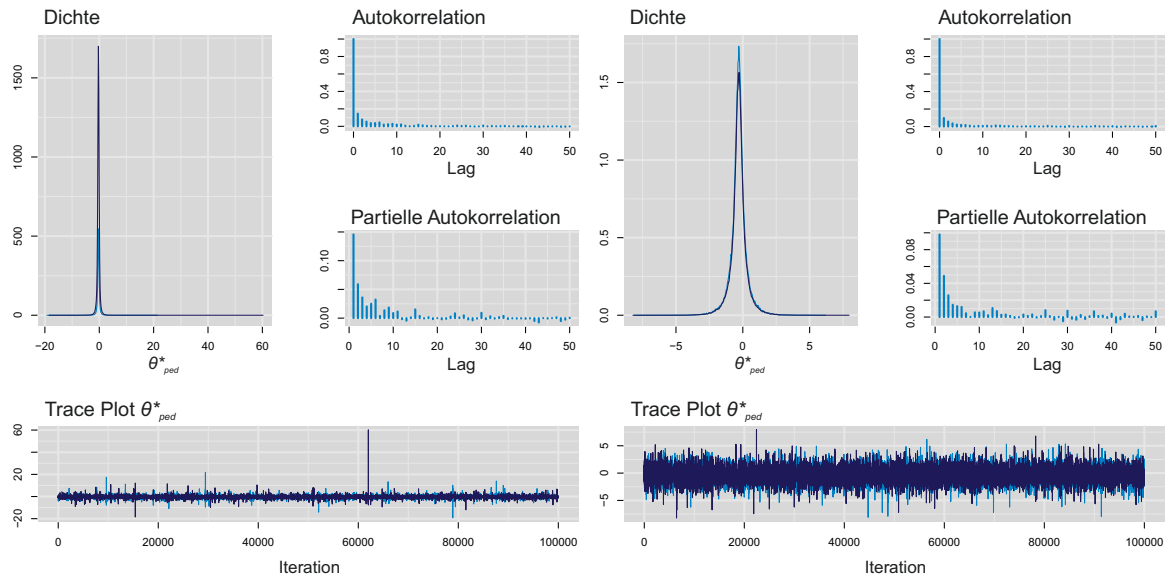


Abb. 4.8: MCMC-Diagnostiken für θ_{ped}^* im Full-Modell der FeNO-Studien.

Varianz τ^2 führt. Obwohl beide Markovketten fast identisch verlaufen, ist es fraglich, ob die richtige gemeinsame stationäre Verteilung erreicht worden ist. Bei der Gleichverteilung können sich die wenigen Studienbeobachtungen nicht gegen die a-priori Verteilung in der a-posteriori Verteilung durchsetzen. Konsequenterweise überdecken die Samples den gesamten zugelassenen Wertebereich $[0, 100]$.

Die unterschiedlichen Schätzungen bestätigen sich in Tabelle 4.5. Während die Standardabweichung bei der Half-Normal-Verteilung mit $\tau = 0.3986$ und bei der Half-Cauchy-Verteilung mit $\tau = 0.4022$ im Mittel fast identisch geschätzt wird, weicht die Schätzung bei der Inverse-Gamma-Verteilung mit $\tau = 0.2561$ stark nach unten ab. Im Gegensatz dazu fällt die Schätzung $\tau = 0.6752$ durch die schwach-informative Gleichverteilung sehr konservativ aus. Dies ist jedoch weit weniger nachteilig als eine Überschätzung der Homogenität der historischen Daten. In der vorliegenden Studiensituation ergibt sich somit die Hypothese, dass die Interstudienstandardabweichung τ bei Verwendung von $\Gamma^{-1}(0.001, 0.001)$ als hyper a-priori Verteilung für τ^2 unterschätzt wird. Weil wir aber das wahre τ nicht kennen, könnte es allerdings auch sein, dass die Schätzung durch die Inverse-Gamma-Verteilung richtig ist und die aller anderen Verteilungen zu groß und damit zu konservativ. Da wir die Hypothese in diesem Beispiel nicht belegen können, werden wir sie später in den Simulationsszenarien bei bekanntem wahren τ überprüfen (vgl. Kap. 5).

Betrachten wir nun in Tabelle 4.5 die a-posteriori prädiktiven Behandlungseffekte θ_{ped}^* und θ_{ad}^* . Die Punktschätzer, d. h. die a-posteriori prädiktiven Erwartungswerte, sind bei allen Verteilungen fast identisch mit $-0.310 \leq \theta_{ped}^* \leq -0.301$ bzw. $-0.177 \leq \theta_{ad}^* \leq -0.171$. Was für die Erwartungswerte gilt, ändert sich jedoch bei den Varianzen. Wir erkennen auffällige Unterschiede bei den verschiedenen a-priori Verteilungen. Die Half-Normal- und die Half-Cauchy-Verteilung schätzen die Standardabweichung ungefähr gleich mit $V_{ped, \tau, \mu_{shift}}^{1/2} = 0.637$ bzw. $V_{ped, \tau, \mu_{shift}}^{1/2} = 0.676$. Die Verkleinerung der Varianz im getrimmten Modell liegt in einem tolerierbaren Bereich. Die Gleichverteilung und die Inverse-Gamma-Verteilung liefern wie in den Beispielen zuvor abweichende Ergebnisse. Die Gleichverteilung $U(0, 100)$ ist für diese Studien zu schwach-informativ bzw. zu vage. Deswegen reichen die wenigen Beobachtungen nicht aus, um sich in der a-posteriori prädiktiven Verteilung gegen die a-priori Gleichverteilung durchzusetzen. Mit $V_{ped, \tau, \mu_{shift}}^{1/2} = 1.785$ wird die Standardabweichung mehr als doppelt so hoch geschätzt wie die durch die Half-Normal- oder die Half-Cauchy-Verteilung. Wie wir im Traceplot 4.8c für θ_{ped}^* sehen, werden immer wieder einzelne sehr hohe oder sehr niedrige Samples generiert. Diese Werte sind völlig unrealistisch und werden in der klinischen Realität nie als tatsächliche Studieneffekte auftreten. Behandlungsunterschiede dieses Ausmaßes zwi-

schen den beiden Behandlungsgruppen kann es nicht geben, da die Therapieform in beiden Armen gleich ist. Lediglich die Einstellung der Therapie unterscheidet sich. Die Extremwerte von θ_{ped}^* treten immer dann auf, wenn auch τ (vgl. Abb. 4.7c) unrealistisch groß geschätzt wird. Durch die Trimming-Methode werden diese Ausreißer entfernt und die Standardabweichung auf 1.146 reduziert. Die Schätzung von θ_{ped}^* ist demnach nicht robust. Obwohl bei der Inverse-Gamma-Verteilung mit fortschreitenden Iterationen ebenfalls immer wieder Ausreißer generiert werden, liefert sie mit $V_{ped,\tau,\mu_{shift}}^{1/2} = 0.541$ dennoch eine kleinere Standardabweichung für θ_{ped}^* als alle anderen a-priori Verteilungen. Wenn wir die Ausreißer durch das getrimmte Modell entfernen, verringert sich folglich die Streuung, hier auf $V_{ped,\tau,\mu_{shift}}^{1/2} = 0.455$. Diese Resultate spiegeln sich direkt im Gewichtungsfaktor der Präzisionen wieder. Die Gleichverteilung sagt das studienspezifische Log Odds Ratio einer neuen Kinderstudie nur sehr vage vorher, sodass die Präzision auf $\omega_{ped} = 0.6\%$ herabgewichtet wird. Im Gegensatz dazu beträgt der Gewichtungsfaktor bei der Inverse-Gamma-Verteilung mit der kleinsten prädiktiven Standardabweichung $\omega_{ped} = 6.9\%$. Bei der $HN(0,1)$ - und der $HC(0,1)$ -Verteilung liegt er zwischen diesen beiden Extrema mit $\omega_{ped} = 4.9\%$ bzw. $\omega_{ped} = 4.3\%$. Die daraus resultierenden Precision Weight Effective Sample Sizes zeigen die Gewichtung etwas deutlicher. Die Fallzahlen schwanken zwischen $N_{ped,eff} = 6$ bei der Gleichverteilung bis zu $N_{ped,eff} = 67$ bei der Inverse-Gamma-Verteilung. Damit wird durch die $\Gamma^{-1}(0.001,0.001)$ -Verteilung die Anzahl mehr als 1.39 mal so hoch geschätzt wie durch die $HN(0,1)$ - oder $HC(0,1)$ -Verteilung. Das Trimmen verstärkt diesen Effekt zusätzlich. Wir erreichen 94 Patienten, was daran liegt, dass die relativ geringe Standardabweichung durch das Trimmen nochmals verkleinert wird. Da der wahre Behandlungseffekt und die Varianz unbekannt sind, können wir in diesem Beispiel nicht beantworten, welcher Gewichtungsfaktor bzw. welche Precision Weight Effective Sample Size richtig ist.

Die bereits beschriebenen Differenzen sehen wir auch bei den Schätzungen für eine neue Studie an Erwachsenen (vgl. Tab. 4.5). Die Parameter θ_{ad}^* , $V_{ad,\tau,\mu_{shift}}^{1/2}$ und $N_{ad,eff}$ unterscheiden sich wie bei den Kindern hinsichtlich der verwendeten a-priori Verteilung. Die Anzahl der historischen Patienten $N_{hist} = 979$ wird auf $\omega_{ad} = 5.6\%$ bis minimal $\omega_{ad} = 0.55\%$ für eine neue Erwachsenenstudie herabgewichtet. Damit ist der Wertverlust der historischen Information hoch. Wie wir an diesem Beispiel deutlich sehen, ist insgesamt nicht allein die Gesamtfallzahl der historischen Studien entscheidend für die Stärke der Informationskraft, sondern die Anzahl der Studien und deren Homogenität. Da in den beiden Kollektiven jeweils nur 3 bzw. 2 frühere Studien verfügbar sind, ist die Vorhersagekraft für eine neue Studie äußerst gering.

FeNO-Studien	A-priori Verteilung für τ							
	$\Gamma^{-\frac{1}{2}}(0.001, 0.001)$		$HN(0, 1)$		$U(0, 100)$		$HC(0, 1)$	
	α -trim		α -trim		α -trim		α -trim	
N_{hist}	979		979		979		979	
$\mathbb{E}(\mu_{shift} y)$	-0.124	-0.123	-0.136	-0.136	-0.142	-0.142	-0.138	-0.138
$\mathbb{E}(\tau y)$	0.2561	0.2509	0.3986	0.3962	0.6752	0.6462	0.4022	0.3976
Geplante Kinderstudie								
$\mathbb{E}(\theta_{ped}^* y)$	-0.301	-0.302	-0.308	-0.308	-0.310	-0.308	-0.308	-0.308
$V_{ped, \tau, \mu_{shift}}^{1/2}$	0.541	0.455	0.637	0.608	1.785	1.146	0.676	0.625
$V_{ped, \tau \equiv 0, \mu_{shift_0}}^{1/2}$	0.1416		0.1416		0.1416		0.1416	
ω_{ped}	0.0685	0.0968	0.0491	0.0543	0.0063	0.0153	0.0433	0.0513
$N_{ped, eff}$	67	94	48	53	6	14	43	50
Geplante Erwachsenenstudie								
$\mathbb{E}(\theta_{ad}^* y)$	-0.177	-0.177	-0.172	-0.172	-0.171	-0.167	-0.172	-0.172
$V_{ad, \tau, \mu_{shift}}^{1/2}$	0.599	0.578	0.713	0.708	1.913	1.461	0.758	0.741
$V_{ad, \tau \equiv 0, \mu_{shift_0}}^{1/2}$	0.1416		0.1416		0.1416		0.1416	
ω_{ad}	0.0559	0.0590	0.0394	0.0396	0.0055	0.0087	0.0349	0.0365
$N_{ad, eff}$	54	58	38	39	5	9	34	35

Tab. 4.5: FeNO-Studien: Ergebnisse der meta-analytisch prädiktiven Methode für unterschiedliche a-priori Verteilungen für τ .

A-posteriori Mittelwerte und Standardabweichungen der wichtigsten Parameter, historische Fallzahl N_{hist} , Precision Weight Effective Sample Sizes $N_{ped, eff}$ und $N_{ad, eff}$, Gewichtungsfaktoren ω_{ped} und ω_{ad} (MCMC-Simulationen: Chains=2, Iterationen=110 000, Burn-In=10 000. Getrimmtes Modell: $\alpha = 0.001$, gültige Werte=2 · 99800).

Die FeNO-Studien werfen daher einige neue Fragen bezüglich der Anwendung des prädiktiven Bayes-Verfahren auf. Unter anderem die Frage, wie viele Studien wir innerhalb eines Kollektivs mindestens benötigen, um eine sinnvolle und richtige Einschätzung des Informationsgehalts der historischen Daten sowie robuste Parameterschätzungen zu erhalten. Außerdem, welche hyper a-priori Verteilung für τ bzw. τ^2 sich in solchen Situationen am besten eignet. Die Antworten geben wir mit Hilfe der Simulationsstudie in Kapitel 5.

Validierung. Als Nächstes validieren wir die ermittelten Precision Weight Effective Sample Sizes für dieses Beispiel (vgl. Kap. 3.3.3). Hierzu planen wir zunächst eine neue Studie an Kindern im Parallelgruppendesign. Wir überprüfen die einseitige Hypothese H_0 (3.17) zum Nachweis der Überlegenheit der Asthmatherapieeinstellung durch FeNO gegenüber der Standardbehandlung ohne Therapieadjustierung. Als Planungsgrundlage für die Ereignisraten der

neuen Studie wählen wir die gepoolten ungewichteten Ereignisraten der beobachteten Kinderstudien, also $p_T = 118/393 = 30.03\%$ und $p_C = 140/389 = 35.99\%$. Damit erhalten wir einen erwarteten Behandlungseffekt von $\text{LogOR} = -0.270$ ($OR = 0.763$), der damit etwas schwächer ist als der vorhergesagte Therapieeffekt durch die historischen Studien. Die geplante Power soll 80% betragen und das Signifikanzniveau $\alpha = 0.025$. Für die Fallzahlschätzung verwenden wir das Normal-Normal-Modell (Kap. 3.18) ohne informative a-priori Funktionen, wobei die Power simuliert wird. Als Resultat erhalten wir die Referenzfallzahl von $N_{ref} = 2 \cdot 975 = 1950$ Patienten, die wir in diesem Design für die Ablehnung der Nullhypothese benötigen. Anschließend untersuchen wir die Power, wenn wir die a-posteriori prädiktiven Erwartungswerte und Varianzen als informative Priorparameter nutzen und die Referenzfallzahl N_{ref} um die Precision Weight Effective Sample Size $N_{ped,eff}$ reduzieren. Die Ergebnisse sind in Tabelle 4.6 dargestellt. Da der nachzuweisende Behandlungseffekt ($\text{LogOR} = -0.270$) etwas geringer ist als die a-posteriori Erwartungswerte ($\text{LogOR} \leq -0.301$), ist zu erwarten, dass die Power etwas über 80% liegt. Diese Erwartung bestätigt sich in allen Fällen. Die bayesianische Power liegt knapp über 80% egal welche hyper a-priori Verteilung für das volle Modell angewandt wird. Wegen der hohen benötigten Fallzahl ist ein möglicher Powerverlust durch eine Reduzierung der Fallzahl in der Größenordnung von $N_{ped,eff}$ relativ gering. Wenn wir beispielsweise 100 Patienten weniger rekrutieren und kein Vorwissen einbringen, beträgt die Power immer bei 77.94%, d. h. wir verlieren nur 2% Power.

Die heuristische Umrechnung der a-posteriori prädiktiven Präzision in die Precision Weight Effective Sample Size scheint auch in diesem Beispiel soweit korrekt zu funktionieren. Natürlich ist Einhaltung der Power in der neuen Studie immer abhängig von den getroffenen Annahmen wie dem Therapieeffekt, dem Signifikanzniveau etc. Der Einfluss der informativen a-priori Parameter auf die a-posteriori Verteilung der neuen Studie ändert sich je nach Design. Da hier der angenommene Behandlungseffekt nicht exakt dem vorhergesagten Effekt entspricht, kann die Precision Weight Effective Sample Size in dieser Situation nicht genau dem Wert der informativen a-priori Parametern entsprechen. Die Precision Weight Effective Sample Size bleibt für jedes Szenario gleich, da sie bereits vor Durchführung der Studie feststeht. Sie kann somit nie exakt mit der reellen einzusparenden Anzahl Patienten, der Prior Effective Sample Size, übereinstimmen.

Prior Effective Sample Size. Aus diesem Grund bestimmen wir auch für die FeNO-Studien $N_{ped,prior}$ für eine neue Kinderstudie und $N_{ad,prior}$ für eine neue Erwachsenenstudie. Wir berechnen die Power mit den Gleichungen in Kapitel 3.4. Zum Nachweis eines Log Odds

FeNO-Studien		Normal-Normal-Modell mit nicht-informativer und informativer a-priori Verteilung									
		Hyper a-priori Verteilung für τ bzw. τ^2 im vollen Modell									
		$\tau^2 \sim \Gamma(0.001, 0.001)$	$\tau \sim HN(0, 1)$	$\tau \sim U(0, 100)$	$\tau \sim HC(0, 1)$						
		α -trim	α -trim	α -trim	α -trim						
Angenommene		<div style="display: flex; justify-content: space-between;"> interventionelle Therapie $p_T = 118/393 = 30.03\%$ </div>									
Ereignisraten		<div style="display: flex; justify-content: space-between;"> konventionelle Therapie $p_C = 140/389 = 35.99\%$ </div>									
A-priori Verteilung	μ_0	0	0	-0.301	-0.302	-0.308	-0.308	-0.310	-0.310	-0.308	-0.308
$\theta \mu_0, \tau_0^2 \sim N(\mu_0, \tau_0^2)$	τ_0	∞	∞	0.541	0.455	0.637	0.608	1.785	1.146	0.676	0.625
Fallzahl der neuen pädiatrischen Studie		2-975 =1950	2-925 =1850	1950 -67 =1883	1950 -94 =1856	1950 -48 =1902	1950 -53 =1897	1950-6 =1944	1950 -14 =1936	1950 -42 =1908	1950 -50 =1900
Bayesianische Power		80.0%	77.94%	80.59%	80.91%	80.41%	80.56%	80.12%	80.11%	80.38%	80.50%

Tab. 4.6: FeNO-Studien: Poweranalyse zur Validierung der errechneten Precision Weight Effective Sample Size.

Normal-Normal-Modell mit nicht-informativen und informativen a-priori Verteilungen für den Studieneffekt der neuen Kinderstudie. Die informativen Parameter sind der a-posteriori prädiktive Erwartungswert $\mathbb{E}(\theta_{ped}^* | y)$ und die a-posteriori prädiktive Varianz $V_{ped, \tau, \mu_{shif, t}}$ aus dem jeweiligen Modell der meta-analytisch prädiktiven Methode. Bei den informativen a-priori Verteilungen ist die Fallzahl um $N_{ped, eff}$ reduziert.

Ratios mit $\text{LogOR} = -0.270 < 0$ in der im vorherigen Abschnitt beschriebenen Kinderstudie, benötigen wir ohne informative Prior die leicht abgeänderte Fallzahl von $N_{ped,ref} = 1956$. Der angenommene Therapieeffekt ist weniger stark als der beobachtete historische Behandlungseffekt. Dadurch bekommen die informativen a-priori Parameter mehr Einfluss, was sich konsequenterweise in höheren Prior Effective Sample Sizes $N_{ped,prior}$ im Vergleich zu $N_{ped,eff}$ äußert. Insgesamt können durch die Extrapolation der historischen Information je nach hyper a-priori Verteilung des vollen Modells zwischen 9 und 95 Patienten weniger rekrutiert werden (vgl. Tab. 4.7). Auch hier bedingen die großen Unterschiede der Präzisionen sehr heterogene Fallzahlen. Angesichts des Wunsches nach Einhaltung der Power, stehen wir vor der schwierigen Entscheidung, um welche Fallzahl wir die neue Studie tatsächlich reduzieren können. Vertrauen wir den Schätzungen durch die Inverse-Gamma-Verteilung, können wir 95 Kinderpatienten einsparen. Sind wir uns sehr unsicher über den wahren Effekt der historischen Studien und gehen vom Worst Case aus, benutzen wir die Schätzungen der Gleichverteilung. Dadurch können wir aber nur 9 Patienten einsparen. In Anbetracht der Tatsache, dass wir für die neue Studie 1956 Patienten benötigen, sind beide Einsparungen sehr gering. Die relative Ersparnis beträgt 0.05% bis 5%. Daher würden wir für die Planung und Auswertung der neuen Studie auf die Verwendung der a-priori Information der historischen Studien verzichten und ein komplett neues Patientenkollektiv rekrutieren.

Zur Bestimmung der Prior Effective Sample $N_{ad,prior}$ planen wir zusätzlich eine neue Studie an Erwachsenen. Diese planen wir unter Annahme des historischen Behandlungseffekts $\text{logOR} = -0.172$, den wir bei Anwendung der Half-Normal- und Half-Cauchy-Verteilung als a-posteriori prädiktiven Erwartungswerte erhalten haben. Zusätzlich unterstellen wir die gemeinsame beobachtete Ereignisrate der Erwachsenen von $p_{TC} = (26 + 30)/(98 + 89) = 30\%$. Um den geringen Behandlungseffekt nachzuweisen, benötigen wir $N_{ad,ref} = 5074$ neue Patienten. Für eine Power von 80% und Signifikanzniveau von 2.5% können wir auch bei dieser speziellen Erwachsenenstudie tatsächlich mehr Patienten $N_{ad,prior}$ einsparen, als von $N_{ad,eff}$ grob vorhergesagt (vgl. Tab. 4.5, Tab. 4.7). Die Reduktion liegt zwischen 8 bis 73 Patienten. Die relative Verringerung beträgt aber nur 0.2% bis maximal 1.5%. Damit lohnt es sich auch für die neue Erwachsenenstudie nicht, die historischen informativen a-priori Parameter in der Auswertung zu verwenden.

Das Trimmen der Markovketten im vollen Modell führt bei der neuen Kinderstudie genauso wie bei der neuen Erwachsenenstudie zu stark abweichenden virtuellen Patientenzahlen $N_{ad,ped}$ und $N_{ad,prior}$. Die extremsten Unterschiede zwischen $N_{ad,ped}$ bei Verwendung der getrimmten informativen Parameter im Vergleich zu $N_{ad,ped}$ unter Anwendung der ungetrimm-

FeNO-Studien	Normal-Normal-Modell mit informativer a-priori Verteilung							
	Hyper a-priori Verteilung für τ bzw. τ^2 im vollen Modell							
	$\Gamma^{-\frac{1}{2}}(0.001, 0.001)$		$HN(0, 1)$		$U(0, 100)$		$HC(0, 1)$	
	α -trim		α -trim		α -trim		α -trim	
Geplante Kinderstudie	$p_T = 118/393 = 30.03\%$				$p_C = 140/389 = 35.99\%$			
	logOR = -0.2703							
$N_{ped,ref}$	1956							
$\mu_{ped,0}$	-0.301	-0.302	-0.308	-0.308	-0.310	-0.308	-0.308	-0.308
$\tau_{ped,0}$	0.541	0.455	0.637	0.608	1.785	1.146	0.676	0.625
$N_{ped,prior}$	95	136	71	78	9	22	63	74
$N_{ped,prior}/N_{ped,ref}$	0.049	0.070	0.037	0.040	0.005	0.011	0.032	0.038
$N_{ped,prior}/N_{hist}$	0.097	0.139	0.073	0.080	0.009	0.022	0.064	0.076
Geplante Erwachsenenstudie	$p_T = 28.14\%$				$p_C = 31.75\%$			
	logOR = -0.172							
$N_{ad,ref}$	5074							
$\mu_{ad,0}$	-0.177	-0.177	-0.172	-0.172	-0.171	-0.167	-0.172	-0.172
$\tau_{ad,0}$	0.599	0.578	0.713	0.708	1.913	1.461	0.758	0.741
$N_{ad,prior}$	73	79	50	51	8	12	44	46
$N_{ad,prior}/N_{ad,ref}$	0.015	0.016	0.010	0.010	0.002	0.002	0.009	0.009
$N_{ad,prior}/N_{hist}$	0.075	0.081	0.051	0.052	0.008	0.012	0.045	0.047

Tab. 4.7: Heparin-Studien: Bestimmung der Prior Effective Sample Size N_{prior} für eine neue Studie unter Verwendung der Parameterschätzungen des vollen Modells für unterschiedliche a-priori Verteilungen für τ .

Informative a-priori Parameter der neuen Studie $\mu_{ped,0}$, $\mu_{ad,0}$, $\tau_{ped,0}$ und $\tau_{ad,0}$, benötigte Fallzahl mit nicht-informativen a-priori Verteilungen $N_{ped,ref}$ und $N_{ad,ref}$, Prior Effective Sample Sizes $N_{ped,prior}$ und $N_{ad,prior}$. Normal-Normal-Modell mit nicht-informativen und informativen a-priori Verteilungen für den Studieneffekt einer neuen Kinder- und Erwachsenenstudie. Die informativen Parameter der Kinderstudie sind der a-posteriori prädiktive Erwartungswert $\mu_{ped,0} = \mathbb{E}(\theta_{ped}^*|y)$ und die a-posteriori prädiktive Varianz $\tau_{ped,0}^2 = V_{ped,\tau,\mu_{shif t}}$ aus dem vollen Modell der meta-analytisch prädiktiven Methode. Für die Erwachsenenstudie sind die informativen a-priori Parameter analog $\mu_{ad,0} = \mathbb{E}(\theta_{ad}^*|y)$ und $\tau_{ad,0}^2 = V_{ad,\tau,\mu_{shif t}}$.

ten informativen Parameter treten bei der Inverse-Gamma-Verteilung auf. Positiv verhält sich die Half-Normal-Verteilung, bei der die Unterschiede der getrimmten Fallzahl sowohl für $N_{ad,ped}$ als $N_{ad,prior}$ am geringsten ausfallen. Sie bewirkt in beiden Fällen die robustesten Schätzungen, gleich gefolgt von der Half-Cauchy-Verteilung. Allerdings sind beide für die hier untersuchte Zielgröße auch die informativsten hyper a-priori Verteilungen.

5 Simulationsstudie

Die Anwendungsbeispiele haben einige neue Fragen aufgeworfen, die wir dort allerdings nicht definitiv beantworten konnten. Dies lag daran, dass wir den wahren Behandlungseffekt und die wahre Streuung der Parameter zwischen den Studien nicht kannten. Dies spiegelt allerdings die Realität wieder. Für eine neue Heparin-Kinderstudie und eine neue FeNO-Kinderstudie benötigten wir in beiden Fällen weit über 1000 Patienten und für eine neue Erwachsenenstudie sogar mehr als 5000 Patienten. Bei den Kinderstudien wird die bayesianische Power unter Verwendung informativer Prior und Reduzierung der Fallzahl um $N_{ped,eff}$ eingehalten. Ist dies aber immer noch so, wenn die Fallzahl der neuen Studie in einer für die praktische Umsetzung realistischeren Größenordnung liegt? Wird dann die Power unter Anwendung der Precision Weight Effective Sample Size ebenfalls eingehalten? Um diese Fragen zu beantworten, führen wir in diesem Kapitel eine Simulationsstudie durch. Hierbei werden wir verschiedene vorgegebene Szenarien für die historischen Studien mit unterschiedlichen Fallzahlen und Effekten betrachten. Mit der Simulationsstudie wollen wir den Fragen aus der Einleitung (Kap. 1) genauer nachgehen. Diese waren:

- Entspricht die Precision Weight Effective Sample Size der a-priori Information über den zukünftigen Behandlungseffekts einer zukünftigen Studie?
- Wird die bayesianische Power in einer neuen Studie unter Verwendung der Precision Weight Effective Sample Size eingehalten?
- Stimmt die Precision Weight Effective Sample Size mit der Prior Effective Sample Size überein?
- Wie viele historische Studien sind nötig, um den studienspezifischen Effekt einer zukünftigen Studie stabil und präzise vorherzusagen?
- Welche hyper a-priori Verteilung für die Interstudienvarianz ist bei geringer Anzahl an Vorstudien geeignet?

In den Simulationen betrachten wir den meta-analytisch prädiktiven Ansatz zur Vereinfachung nur für ein Kollektiv. Ohne Beschränkung der Allgemeinheit nehmen wir an, dass es sich hierbei um ein Kollektiv von pädiatrischen Patienten handelt. Die Heterogenität der Studien setzen wir als moderat voraus. Wir planen eine neue Studie im Parallelgruppendesign, die die Überlegenheit der innovativen Behandlung gegenüber der Standardtherapie zeigen

soll. Dafür betrachten wir einen binären Endpunkt mit dem logarithmierten Odds Ratio als Zielgröße. Wir untersuchen die einseitigen Hypothesen (vgl. 3.17)

$$H_0 : OR = \frac{Odds(p_T)}{Odds(p_C)} \geq 1 \quad \text{versus} \quad H_1 : OR = \frac{Odds(p_T)}{Odds(p_C)} < 1, \quad (5.1)$$

wobei wie zuvor p_T die Ereignisrate unter der innovativen Therapie und p_C die Ereignisrate unter der Standardbehandlung sei. Das bayesianische Signifikanzniveau (vgl. 3.20) sei $\alpha = 0.025$ und die bayesianische Power 80%. Die Simulationsstudie gliedert sich in die folgenden Schritte: Planung der neuen Studie ohne Vorwissen, Generierung der historischen Studiendaten für die bayesianische Metaanalyse, Durchführung der meta-analytisch prädiktiven Methode sowie verschiedene Poweranalysen für die neue Studie mit informativer a-priori Verteilung und reduzierter Fallzahl.

5.1 Settings

Zuerst planen wir die neue Studie und ermitteln mit dem Normal-Normal-Modell und den Powersimulationen aus Kapitel 3.3.3 die benötigte Fallzahl ohne die Nutzung von Vorwissen. Wir simulieren die Power aus Gründen der Konsistenz, da wir nach dem gleichen Prinzip auch die Ergebnisse der historischen Studien generieren.

Bemerkung 5.1. Die nachstehende Strategie lässt sich insgesamt wie folgt interpretieren: Wir planen die neue Studie bezüglich der Annahme, dass wir den wahren Behandlungseffekt kennen. Dafür berücksichtigen wir jedoch die Schätzer wie den a-posteriori prädiktiven Erwartungswert und die Varianz aus den generierten historischen Daten nicht. Im klinischen Alltag wäre dies so, als wenn die Studienplaner für die Planung, aus welchen Gründen auch immer, nur ihr Expertenwissen über das mögliche Outcome der neuen Studie nutzen und dieses über die Vorhersagen aus den historischen Daten stellen. Da sie sich aber eventuell doch nicht ganz sicher über die Richtigkeit ihrer Annahmen sind, wollen sie die historischen Daten aber auch nicht komplett ignorieren. Darum berücksichtigen sie das historische Vorwissen in der Auswertung der neuen Studie. Die a-posteriori prädiktiven Parameter gehen somit als informative a-priori Parameter in die Auswertung der neuen Studie ein, die aber auf den wahren Behandlungseffekt gepowert wurde. Wenn die Annahmen stark von den historischen prädiktiven Parametern abweichen, führt dies dazu, dass das Powerniveau nicht mehr einge-

halten wird. Ist der informative a-priori Erwartungswert kleiner oder größer als der für die Planung zugrunde gelegte Behandlungseffekt und gleichzeitig die a-priori Präzision sehr groß, ist die neue Studie unter- oder überpowert. Allerdings ist dies im klinischen Alltag gängig, denn der angenommene Therapieeffekt weicht oft von der tatsächlichen Beobachtung ab.

Im Folgenden betrachten wir zwei unterschiedliche Szenarien. Eine Übersicht der Annahmen und Einstellungen findet sich in Tabelle 5.1.

5.1.1. Szenario A

Planung der neuen Studie. In Szenario A nehmen wir an, dass der wahre Behandlungseffekt der Grundgesamtheit einem Odds Ratio von $OR_0 = 0.5$ entspricht. Das wahre Log Odds Ratio beträgt demnach $\mu_0 = -0.693$. Als gemittelte Ereigniswahrscheinlichkeit p_{TC} über beide Behandlungsgruppen wählen wir $p_{TC} = 5/12 = 41.67\%$. Mit dem gewählten Odds Ratio ergibt sich damit für die Ereignisrate unter der innovativen Therapie $p_T = 1/3 = 33.3\%$ und für die Ereignisrate unter der Standardtherapie $p_C = 50\%$. Weiter nehmen wir an, dass die Heterogenität der Studien moderat ist. Da im Intervall $\mu \pm 1.96\tau$ ca. 95% der Werte von θ liegen, ergibt sich damit das Verhältnis des 97.5% OR zum 2.5% OR von $\exp(3.92\tau)$. Dieser Ausdruck repräsentiert einen verhältnismäßigen „Range“ der Odds Ratios zwischen den Studien [vgl. 62, S. 169]. *Smith et al.* [59] schreiben, dass es sehr unwahrscheinlich ist, im Kontext von Metaanalysen eine größere Streuung der Odds Ratios als $\exp(3.92\tau) = 10$, also $\tau = 0.59$, zu beobachten. *Spiegelhalter et al.* [62] geben in diesem Zusammenhang folgende Interpretation der Werte für τ : Werte von 0.1 bis 0.5 sind annehmbar, von 0.5 bis 1.0 als ziemlich hoch einzustufen, und Werte über 1.0 stehen für eine extreme Heterogenität. Für die Annahme einer moderaten Heterogenität setzen wir deshalb die wahre Interstudienstandardabweichung auf $\tau_0 = 0.5$. Dies entspricht einem „Range“ der Odds Ratios von $\exp(3.92\tau) = 7.1$. Für die Planung der neuen Studie benötigen wir die Standardabweichung nicht, jedoch für die Generierung der studienspezifischen Effekte der historischen Studien.

Das Normal-Normal-Modell mit nicht-informativer a-priori Verteilung, also ohne Einbringung von Vorwissen, und die zugehörige Poweranalyse liefern die benötigte Fallzahl von $N_{ref} = 2 \cdot 135 = 270$ auswertbaren Patienten in der neuen Studie. Diese Fallzahl entspricht einer Größe, die für die praktische Umsetzung geeignet ist und die tatsächliche Durchführung einer neuen klinischen Studie realisierbar macht.

Historische Daten. Die historische Gesamtfallzahl aller Studien setzen wir in Szenario A fest auf $N_{hist} = 600$. Wir variieren die Anzahlen der historischen Studien und betrachten jeweils 3, 4, 5 und 6 historische Studien als Grundlage pro Metaanalyse. Die Fallzahl in jeder Studie ist dabei gleich. Demnach ergeben sich die Fallzahlen $N = 200, 150, 120, 100$ der einzelnen historischen Studien. Desweiteren sei das Gruppenverhältnis zwischen Behandlungs- und Kontrollgruppe gleich, sodass sich in jedem Studienarm gleich viele Patienten befinden.

5.1.2. Szenario B

Planung der neuen Studie. In Szenario B unterstellen wir einen etwas kleineren Behandlungseffekt, während wir aber von einer größeren Anzahl historischer Patienten ausgehen. Wir nehmen an, dass der wahre Behandlungseffekt einem Odds Ratio von $OR_0 = 0.7$ entspricht. Demnach ist das wahre Log Odds Ratio $\mu_0 = -0.357$. Weiter setzen wir eine gemittelte Ereignisrate über beide Behandlungsgruppen von $p_{TC} = 35.91\%$ voraus. Daraus ergibt sich für die erwartete Ereignisrate unter der innovativen Therapie $p_T = 31.82\%$ und für die erwartete Ereigniswahrscheinlichkeit unter der Standardtherapie in der Kontrollgruppe $p_C = 40\%$. Mit der Methodik aus Kapitel 3.3.3 führen wir erneut eine Fallzahlschätzung mit nicht-informativer Startverteilung durch. Wir benötigen unter diesen Annahmen $N_{ref} = 2 \cdot 537 = 1074$ auswertbare Patienten zum Nachweis der Alternativhypothese $H_1 : \mu_0 < 0$.

Historische Daten. In diesem Szenario betrachten wir als historische Datenbasis ebenfalls jeweils 3, 4, 5 und 6 Studien. Als Fallzahl der einzelnen Studien wählen wir $N = 334, 250, 200, 167$ mit jeweils gleichen Gruppenverhältnissen. Demzufolge ist die Gesamtfallzahl $N_{hist} = 1000$ für 4 bzw. 5 Studien und $N_{hist} = 1002$ für 3 und 6 Studien.

5.1.3. Generierung der historischen Studien

Für jedes Szenario und jede Studienanzahl generieren wir 10 000 zufällige historische Studienkonstellationen für die Metaanalysen. Dafür ziehen wir für jede Studie den Gruppenparameter θ_h , das „wahre“ Log Odds Ratio der einzelnen Studie, aus der Normalverteilung: $\theta_h \sim N(\mu_0, \tau_0)$ mit $h \in \{1, \dots, H\}$ und $H \in \{3, 4, 5, 6\}$. Unter der Annahme der vorgegebenen festen gemittelten Ereigniswahrscheinlichkeit p_{TC} erzeugen wir für das studienspezifische Log Odds Ratio eine Kontingenztafel mit insgesamt 10 000 000 Beobachtungen. Über diesen Hilfschritt können wir für jede Studie die studienspezifischen „wahren“ Ereignisraten

	Szenario A	Szenario B
OR_0	0.5	0.7
$\mu_0 := \text{Log}OR_0$	-0.693	-0.357
τ_0	0.5	0.5
p_{TC}	41.67%	35.91%
p_T	33.33%	31.82%
p_C	50.0%	40.0%
N_{ref}	270	1074
N_{hist}	600	1000 (1002)
# Studien	Fallzahl pro Studie	
3	$N = 200$	$N = 334$
4	$N = 150$	$N = 250$
5	$N = 120$	$N = 200$
6	$N = 100$	$N = 167$
# Metaanalysen	10 000	
MCMC-Einstellungen		
Markovketten	2	
Burn-In	10 000	
Iterationen	110 000	
Thinning	10	
Poweranalyse	100 000 Simulationen der neuen Studie pro Metaanalyse	

Tab. 5.1: Übersicht der Simulationsszenarien und Einstellungen.

t_T und t_C bestimmen. Diese Ereignisraten benötigen wir, um schließlich binomialverteilt die beobachteten Ereigniszahlen e_T und e_C der historischen Studie zu generieren

$$e_T \sim \text{Bin}\left(\frac{N}{2}, t_T\right) \quad \text{und} \quad e_C \sim \text{Bin}\left(\frac{N}{2}, t_C\right).$$

Kurzum, wir ziehen also mit den bekannten Hyperparametern μ_0 und τ_0 die Studienparameter θ_h aus der gemeinsamen Normalverteilung. Anschließend nehmen wir an, dass die beobachteten Studienergebnisse zusätzlich wiederum um den „wahren“ Studienparameter streuen und konstruieren demgemäß die beobachteten Ereigniszahlen (vgl. Abb. 2.2).

5.1.4. MCMC-Simulationen

Für jedes Szenario und jede Studienanzahl betrachten wir 10 000 zufällige Konstellationen der historischen Daten. Für jede Studiensituation berechnen wir die in Kapitel 3 beschriebenen hierarchischen Modelle. Wir simulieren das volle Modell sowie das Referenzmodell und führen die Trimming-Methode mit $\alpha = 0.1\%$ durch. Für das MCMC-Sampling des vollen Modells und

des Referenzmodells generieren wir 2 Markovketten mit jeweils 110 000 Iterationsschritten. Als Burn-In werden die ersten 10 000 Updatingschritte für die Auswertung weggelassen. Da zu erwarten ist, dass das Gewicht der simulierten historischen Daten voraussichtlich nicht sehr hoch sein kann, ist mit einer erhöhten Autokorrelation der Iterationsschritte zu rechnen. Aus diesem Grund verwenden wir einen Thinning-Parameter von 10. Das heißt, es wird nur jeder zehnte Schritt einer Markovkette gespeichert. Insgesamt erhalten wir bei jeder Metaanalyse $2 \cdot 10\,000$ gültige Iterationen für die Bestimmung der a-posteriori Verteilungen. Im α -getrimmten Modell beträgt die Anzahl dementsprechend 19 960 Werte. Als Ergebnis der 10 000 einzelnen Studienkonstellation speichern wir den a-posteriori prädiktiven Erwartungswert θ^* sowie die zugehörige Standardabweichung V_τ , den a-posteriori Erwartungswert der Interstudienvariation τ , den Gewichtungsfaktor der Präzisionen ω und die errechnete Precision Weight Effective Sample Size N_{eff} . Das Gleiche gilt für die zugehörigen α -getrimmten Werte. Da wir in beiden Szenarien jeweils den wahren Therapieeffekt μ_0 und dessen Varianz τ_0^2 kennen, können wir außerdem für jede Studiensituation die wahre Precision Weight Effective Sample Size N_{true} berechnen (vgl. 3.16).

5.1.5. Powerbetrachtung

Mit jedem Ergebnis der 10 000 Metaanalysen führen wir eine Poweranalyse durch. Wir benutzen den a-posteriori prädiktiven Erwartungswert und die Standardabweichung von θ^* als informative hyper a-priori Parameter für das Normal-Normal-Modell der neuen Studie. Die Fallzahl der neuen Studie $N_{ref} = 270$ bzw. $N_{ref} = 1074$ reduzieren wir jeweils um N_{eff} . Anschließend simulieren wir 100 000 mal das Ergebnis der zukünftigen Studie unter Einbringung des Vorwissens. Wir notieren die relative Häufigkeit bayesianisch signifikanter Ergebnisse und erhalten somit für jede erzeugte Metaanalysenkonstellation eine bayesianische Power (vgl. Kap 3.3.3).

5.1.6. Berechnung der Prior Effective Sample Size

Zusätzlich berechnen wir die wirkliche Fallzahleinsparung der neuen Studie N_{prior} . Dafür nutzen wir den a-posteriori prädiktiven Erwartungswert $\mathbb{E}(\theta^*|y)$ und die a-posteriori prädiktive Varianz V_τ als informative a-priori Parameter. Die Fallzahl N_{prior} bestimmen wir nach Formel 3.30. Die Prior Effective Sample Size wird unter Konstanthaltung der Power von 80% berechnet. Aus den in Bemerkung 5.1 beschriebenen Gründen, kann es sein, dass die infor-

matischen a-priori Parameter einen anderen Behandlungseffekt suggerieren als in der Planung unterstellt. Darum kann die Fallzahl $N_{ref,inf}$ unter Verwendung der informativen a-priori Parameter, die zum Erreichen einer Power von 80% benötigt wird, nicht nur kleiner sein als die geplante uninformative Fallzahl N_{ref} , sondern diese sogar übersteigen. Nach Definition 3.30 kann damit N_{prior} negativ werden. Ein negatives N_{prior} bedeutet, dass diese Anzahl an Patienten noch zusätzlich zu den ursprünglich benötigten Patienten N_{ref} rekrutiert werden müssen.

5.2 Adjustiertes Simulationssetting zur Validierung der Precision Weight Effective Sample Size

In dem bisherigen Simulationssetting können wir nicht nachweisen, ob einerseits die geschätzte Precision Weight Effective Sample Size N_{eff} der ermittelten a-posteriori prädiktiven Präzision des Studieneffekts entspricht und andererseits, ob sie wirklich mit der Prior Effective Sample Size N_{prior} übereinstimmt. Bei der Berechnung von N_{eff} (vgl. Gleichung (3.7)) werden nur die Präzisionen des vollen Modells und des Referenzmodells berücksichtigt. Der a-posteriori prädiktive Erwartungswert geht nicht in die Evaluierung ein. Außerdem ist N_{ref} unabhängig von den Annahmen der neuen Studie und immer konstant. Falls in der neuen Studie die bei nicht-informativer Planung benötigte Fallzahl N_{ref} um N_{eff} reduziert wird und die historischen informativen Prior verwendet werden, hat dies konsequenterweise Auswirkungen auf die Power. Da wir die historischen Daten für jede Metaanalyse zufällig generieren, kann es vorkommen, dass der a-posteriori prädiktive Erwartungswert nicht mit dem jeweils angenommenen Behandlungseffekt $\mu_0 = -0.693$ bzw. $\mu_0 = -0.357$ der neuen Studie übereinstimmt. Somit ist zu erwarten, dass die Power der neuen Studie nicht 80% betragen kann, wenn ein verzerrter a-posteriori prädiktiver Erwartungswert mit kleiner Varianz als informativer a-priori Therapieeffekt in die Auswertung einfließt.

Im Gegensatz dazu wird für die Bestimmung der Prior Effective Sample Size N_{prior} die Power konstant gehalten, sodass N_{prior} direkt durch die informativen a-priori Parameter beeinflusst wird. Zur Überprüfung der Gleichheit von N_{eff} und N_{prior} müssen wir also eine Situation erzeugen, sodass unter Reduzierung der Fallzahl der neuen Studie um N_{eff} die Power genau 80% betragen müsste. Dies sollte genau dann der Fall sein, wenn der a-posteriori prädiktive Erwartungswert gleich dem angenommenem Effekt ist. Die a-priori Information wird dann

nur durch die a-posteriori prädiktive Varianz V_τ beeinflusst, die auch in die Berechnung von N_{eff} eingeht.

In der klinischen Realität würden wir üblicherweise die neue Studie bezüglich des beobachteten historischen Therapieeffekts planen. Dieser ist für gewöhnlich ein guter Schätzer des unbekanntem wahren Behandlungserfolgs. Das in Bemerkung 5.1 beschriebene Vorgehen, in der die historischen Vorhersagen für die Planung ignoriert werden, ist aus den bereits beschriebenen Gründen nur suboptimal. Um nun zu vermeiden, für jede der 10 000 bereits simulierten Metaanalysedaten auf Basis der beobachteten historischen Effekte eine neue Studie zu planen, bedienen wir uns eines einfachen technischen Tricks. Wir müssten sonst jeweils andere Ereignisraten und Log Odds Ratios für jede neue Studie unterstellen, sodass die Fallzahlschätzungen sehr unterschiedliche Ergebnisse ergeben. Um dies zu umgehen und die neu geplanten Studien vergleichbar zu machen, behelfen wir uns mit einer Adjustierung der historischen Ergebnisse. Wir verschieben die historischen a-posteriori prädiktiven Erwartungswerte $\mathbb{E}(\theta^*|y)$ auf den wahren Effekt. Somit ist der informative a-priori Erwartungswert der neuen Studie in Szenario A bzw. B gleich dem geplanten Studieneffekt, also

$$\text{Szenario: A} \quad \mu_0 \equiv -0.693 \quad \text{und} \quad \text{Szenario: B} \quad \mu_0 \equiv -0.357. \quad (5.2)$$

Dadurch basieren nun die bereits geplanten Studien aus dem vorherigen Abschnitt 5.1, für die wir den wahren Behandlungserfolg als Planungsannahme unterstellt haben, auf dem historischen Effekt. Jede Metaanalyse liefert durch die Adjustierung denselben Punktschätzer, aber unterschiedliche Präzisionen.

In dieser angepassten Situation können wir jetzt herauszufinden, ob die geschätzte Precision Weight Effective Sample Size N_{eff} der informativen a-priori Präzision entspricht. Die Umrechnung der Präzision bzw. des Informationsgehalts der historischen Daten in N_{eff} folgt bekanntermaßen heuristischen Argumenten. Zur Beantwortung der Fragen führen wir eine zusätzliche Poweranalyse für alle 10 000 Metaanalysekonstellationen in Szenario A und B durch. Für die adjustierten Powerbetrachtungen der neuen Studie verwenden wir aus der meta-analytisch prädiktiven Methode nur N_{eff} und die errechnete a-posteriori prädiktive Standardabweichung $V_\tau^{1/2}$. Wir benutzen, wie zuvor im Modell (3.24), $V_\tau^{1/2}$ aus dem vollen Modell als informativen a-priori Hyperparameter der Standardabweichung. Im Gegensatz dazu setzen wir den a-priori Erwartungswert auf den geplanten Effekt (vgl. (5.2)). Anschließend führen wir erneut die neue Studie mit reduzierter Fallzahl durch. Sofern nun die Precision Weight Effective Sample Size genau den zusätzlichen Powergewinn ausgleichen würde, der

durch die Verwendung der informativen a-priori Präzision entsteht, sollte die Power ca. 80% betragen. Falls sie niedriger als 80% ist, wird N_{eff} zu hoch geschätzt. Falls sie 80% übersteigt, ist die Schätzung von N_{eff} eher zu konservativ.

Ferner können wir in dieser adjustierten Situation nachweisen, ob die Precision Weight Effective Sample Size N_{eff} gleich der Prior Effective Sample Size N_{prior} ist. Dazu bestimmen wir die Anzahl der Patienten, um die die neue Studie reduziert werden kann, sodass unter Verwendung der informativen Prior die Power von 80% eingehalten wird. Da der informative a-priori Behandlungseffekt auf den angenommenen Studieneffekt zentriert ist (vgl. (5.2)), wird die Ersparnis an Patienten nur durch die informative a-priori Präzision beeinflusst. Die Prior Effective Sample Size N_{prior} ist damit unabhängig vom informativen a-priori Erwartungswert, da dieser für alle Metaanalysenresultate gleich ist. Somit entspricht N_{prior} genau dem Informationsgehalt über den zukünftigen Studieneffekt, der durch die a-priori Präzision weitergegeben wird. Wenn N_{eff} korrekt vorhergesagt wird, sollten die beiden verschiedenen effektiven Fallzahlen in diesem Setting übereinstimmen.

Bei den adjustierten Untersuchungen haben die historische Studienanzahl und die hyper a-priori Verteilung für τ keinen Einfluss auf die Beantwortung der beiden Fragen. Darum können wir die Ergebnisse für jedes Szenario auch gepoolt darstellen.

5.3 Ergebnisse der Simulationsstudie

In diesem Abschnitt präsentieren wir ausgewählte Ergebnisse, die besonders zur Beantwortung unserer Fragen beitragen. Ausführliche Tabellen und ergänzende Grafiken für alle Szenarien und Studienanzahlen sind im Appendix angegeben. Die verschiedenen Grafiken sind immer identisch aufgebaut (z. B. Abb. B.1). In Teilabbildung (a) sind die Ergebnisse für das Simulationsszenario A dargestellt. In Teilabbildung (b) die Ergebnisse für das Simulationsszenario B. Diese beinhalten wiederum die Ergebnisse aufgeteilt nach Anzahl der vorhandenen historischen Studien (3, 4, 5, 6).

5.3.1. Ergebnisse der nicht adjustierten Simulationen

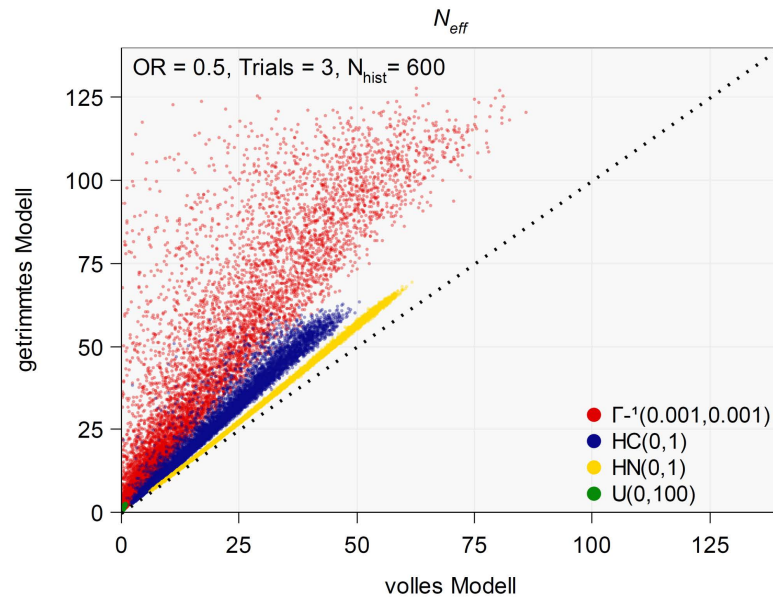
Precision Weight Effective Sample Size N_{eff} . Betrachten wir als ersten Parameter die Precision Weight Effective Sample Size N_{eff} . Am anschaulichsten lassen sich die ermittelten

Anzahlen virtueller Patienten an den Boxplots in Abb. B.1 vergleichen. Dort erkennen wir, dass bei allen Verteilungen mit Erhöhung der historischen Studienanzahl trotz Konstanzhaltung der Gesamtfallzahl N_{hist} , die Fallzahl N_{eff} steigt. Allerdings gibt es große Unterschiede zwischen den hyper a-priori Verteilungen. Während bei der Gleichverteilung die Fallzahl im Median sehr niedrig ist (Szenario A: 0 bis 23 Patienten, Szenario B: 0 bis 25 Patienten) wird sie bei der Inverse-Gamma im Median sehr hoch geschätzt (Szenario A: 9 bis 57 Patienten, Szenario B: 9 bis 119 Patienten). Wir sehen, dass bei der Inverse-Gamma-Verteilung viele sehr hohe Werte entstehen, die die benötigte Fallzahl für die neue Studie $N_{ref} = 270$ in Szenario A fast übersteigen. Die einzelnen Werte sind in Tabelle B.1 und Tabelle B.2 dargestellt. Die Maximalwerte liegen in Szenario A bei $N_{eff} \in (86, 255)$ und in Szenario B bei $N_{eff} \in (118, 394)$. Die Precision Weight Effective Sample Sizes bei der Half-Cauchy- und Half-Normal-Verteilung sind fast identisch und liegen zwischen den Ergebnissen der Gleichverteilung und der Inverse-Gamma-Verteilung. Die wahre Precision Weight Effective Sample Size N_{true} verändert sich nur wenig mit Erhöhung der Studienzahl. Da die wahre Interstudienstandardabweichung τ in beiden Szenarien bekannt ist, wird N_{true} durch die Beobachtungen nur in Form der Intrastudienvarianz beeinflusst (siehe Formel 3.16). Die Intrastudienvarianz ist wiederum nur von den Fallzahlen und Ereignisraten der einzelnen Studien abhängig. Damit kann N_{true} nicht viel variieren. Im Median beträgt N_{true} in Szenario A 48 bis 56 und in Szenario B 51 bis 58. Wenn wir N_{true} als Referenzwert annehmen, können wir erkennen, dass mit der Half-Normal-, Half-Cauchy- und Gleichverteilung dieser Wert unterschätzt wird (Abb. B.1). Bei der Inverse-Gamma-Verteilung ist dies bei 3 historischen Studien ebenfalls noch der Fall. Allerdings ändert sich dies bereits ab 4 Studien. Während der Median von N_{eff} ungefähr N_{true} entspricht, wird N_{eff} in vielen Fällen zu hoch geschätzt. Die übrigen drei Verteilungen halten das Level von N_{true} in etwa ein und führen zu eher konservativen Precision Weight Effective Sample Sizes. Die höheren Fallzahlen kommen durch die zufällige Generierung der historischen Daten zustande. Es können zufällig sehr homogene und sehr heterogene Studienkonstellationen entstehen. Die beobachtete Interstudienvarianz τ^2 kann daher von 0.5^2 abweichen. Eine niedrigere Interstudienvarianz führt dann zu einer viel höheren Precision Weight Effective Sample Size.

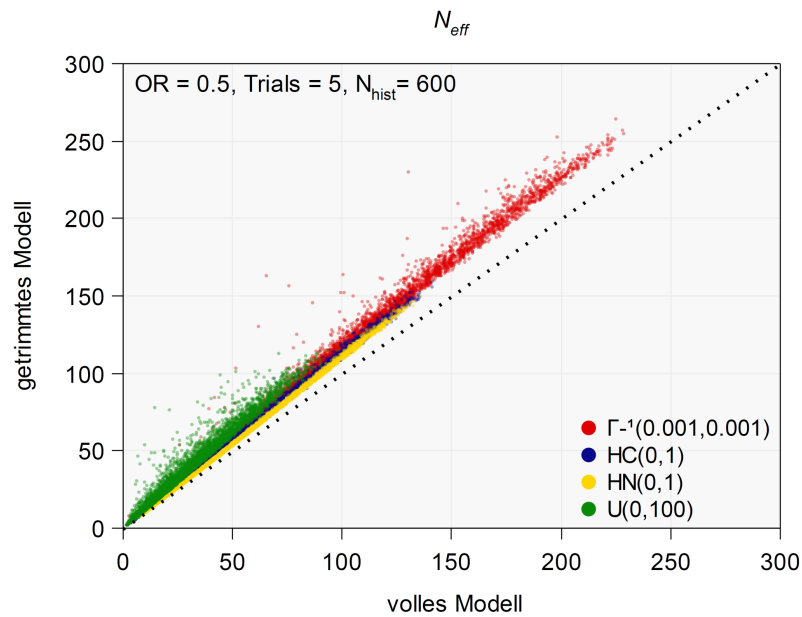
Vergleichen wir N_{eff} im vollen Modell mit N_{eff_trim} im getrimmten Modell, so sehen wir, dass die Unterschiede mit zunehmender Studienanzahl geringer werden. Die Schätzungen der einzelnen Parameter werden durch Erhöhung der Studienzahl genauer. Die hyper a-priori Verteilungen verlieren an Einfluss. Bei den drei Verteilungen $HN(0,1)$, $HC(0,1)$ und $U(0,100)$ gleichen sich die Fallzahlen N_{eff} langsam an. Bei 6 Studien stimmen die Fallzahlen schon

fast überein. Die Inverse-Gamma-Verteilung führt durch ihre Schätzungen in allen Situationen zu tendenziell höheren Precision Weight Effective Sample Sizes als bei den anderen drei Verteilungen. Wenn wir die Situation mit nur 3 Studien betrachten, erkennen wir deutliche Unterschiede zwischen der getrimmten und ungetrimmten Fallzahl. Das Entfernen von 40 extremen Werten in Szenario A führt bei der Inverse-Gamma-Verteilung zu einer 2.14 fachen, bei der Gleichverteilung zu einer 1.93 fachen, bei der Half-Cauchy- zu einer 1.29 fachen und bei der Half-Normal-Verteilung nur zu einer 1.08 fachen medianen Erhöhung der getrimmten Fallzahl gegenüber der ungetrimmten Fallzahl (Tab. B.3a, Tab.B.3b). Die Maxima unterscheiden sich noch viel deutlicher. Positiv fällt hier die Half-Normal-Verteilung auf. Die maximale Erhöhung durch das Trimmen beträgt das 1.7-fache. Sie liefert damit bereits bei 3 Studien die stabilsten Ergebnisse für N_{eff} . Direkt danach folgt die Half-Cauchy-Verteilung, die allerdings vereinzelt sehr extreme Unterschiede zwischen getrimmter und ungetrimmter Fallzahl aufweist (Szenario A, 3 Studien: Max. 40.82 fache Erhöhung). Wir müssen allerdings berücksichtigen, dass es sich hierbei um faktorielle Erhöhungen handeln. Das Maximum 40.82 tritt beispielsweise für $N_{eff} = 0.53$ und $N_{eff_trim} = 21.69$ auf. Hierbei wurde $V_\tau^{1/2}$ von 5.61 auf 0.88 verkleinert. Für 90% der Beobachtungen liegt die Erhöhung aber in einem tolerierbaren Bereich. Bei der Gleichverteilung kann N_{eff} durch das Trimmen in absoluten Zahlen nicht wesentlich erhöht werden, da N_{eff} im Maximum gerade 1 Patient beträgt. Bei mehreren historischen Studien ist das Verhältnis von N_{eff} (getrimmtes Modell/volles Modell) immer noch relativ hoch, wobei die absolute Differenz nur ca. 10 Patienten beträgt. Bei der Inverse-Gamma-Verteilung nähert sich das mediane Verhältnis dem Wert 1.1 an. Die großen Verhältnismaxima, die noch bei 3 und 4 Studien auftreten, verschwinden in Szenario A und B ab 5 Studien. Das maximale Verhältnis beträgt bei 3 Studien 1352.96. Die Fallzahlen sind hier $N_{eff} = 0.07$ und $N_{eff_trim} = 92.98$. Die Standardabweichung $V_\tau^{1/2}$ wird durch das Trimmen von 16.17, einer völlig unrealistischen Schätzung, auf 0.44 gesenkt. Die Bewertung der Daten in diesem extremen Szenario ist mit der $\Gamma^{-1}(0.001, 0.001)$ Startverteilung daher keineswegs robust. In Szenario A bei 6 Studien übersteigt die getrimmte virtuelle Patientenzahl bei der Inverse-Gamma-Verteilung in 20 Fällen die Referenzfallzahl $N_{ref} = 270$ der geplanten Studie. Weil es aufgrund der hohen Bewertung der historischen Ergebnisse nicht mehr notwendig scheint eine neue Studie durchzuführen, berücksichtigen wir diese Resultate für die Poweranalyse nicht.

In Abbildung 5.1 sind exemplarisch die Scatterplots für 3 und 5 Studien für Szenario A dargestellt. Alle anderen Konstellationen sind in den Streupunktdiagrammen B.2 abgebildet. Wir können deutlich die Unterschiede zwischen der ungetrimmten und getrimmten Fallzahl



(a) Vergleich bei 3 Studien.



(b) Vergleich bei 5 Studien.

Abb. 5.1: Scatterplots: Vergleich zwischen N_{eff} im vollen und im getrimmten Modell in Szenario A bei 3 und 5 Studien.

erkennen. Im Idealfall hat das Trimmen des Modells keinen großen Einfluss auf N_{eff} . Die getrimmte virtuelle Fallzahl sollte dann geringfügig größer als die ungetrimmte Fallzahl sein. Die Punkte in den Abbildungen sollten demnach für jede der Verteilungen auf einer gemeinsamen Geraden liegen, die leicht steiler verläuft als die Winkelhalbierende. Dies ist in den

Szenarien bei 5 und 6 historischen Studien der Fall. Die Masse der Punkte formt bei jeder hyper a-priori Verteilung eine Gerade (Abb. 5.1). Ab 5 Studien sind die Residuen deutlich kleiner. Es gibt weniger Ausreißerpunkte, die von der hypothetischen Regressionsgeraden abweichen. Bei der Gleichverteilung und Inverse-Gamma-Verteilung treten diese jedoch noch vergleichsweise häufig auf. Ist die Studienzahl klein, das heißt 3 oder 4 Studien, so sind die Residuen groß. Bei der Inverse-Gamma-Verteilung ist der Einfluss auf N_{eff} durch Anwendung der α -Trimming-Methode am größten. Wie bereits weiter oben gezeigt, wird N_{eff} durch das Trimmen teils wesentlich vergrößert. Die Korrelation zwischen der getrimmten und ungetrimmten Fallzahl ist niedrig. Auch bei der Half-Cauchy-Verteilung ist die Streuung noch sehr hoch. Die Schätzungen sind demnach bei 3 oder 4 Studien für diese Startverteilungen nicht robust. Wir können jedoch bereits einen linearen Trend erkennen. Bei der Gleichverteilung reduzieren sich die Fallzahlen in dieser Darstellung fast auf einen Punkt (grün), da N_{eff} in beiden Modellen nur 0.3 bis 2.8 beträgt. Die Half-Normal-Verteilung performt bei 3 Studien bereits am besten. Die Punkte (gelb) stellen eine klar abgrenzbare Gerade dar. Die Streuung ist äußerst gering. Durch die Erhöhung der Studienanzahl nähern sich die Punkte immer mehr einer eindeutigen Regressiongeraden an. Die Korrelation nimmt zu und die Residuen ab.

Allein durch die Betrachtung der Precision Weight Effective Sample Size N_{eff} können wir bereits davon ausgehen, dass bei 3 Studien keine zuverlässige Bewertung der historischen Daten und Prognose eines zukünftigen Behandlungseffekts möglich sind.

Gewichtungsfaktor des historischen Informationsgehalts ω . Der Gewichtungsfaktor ω ist nicht anderes als N_{eff} multipliziert mit einer Konstanten, hier der historischen Fallzahl $N_{hist} = 600$ bzw. $N_{hist} = 1000$ (1002). Deshalb gelten für ω dieselben Aussagen, die wir im vorherigen Abschnitt für N_{eff} getroffen haben. Die Ergebnisse für den Gewichtungsfaktor sind in Abbildung B.3 und den Tabellen B.4 sowie B.5 zusammengefasst.

A-posteriori prädiktive Erwartungswerte von θ^* . Als Nächstes befassen wir uns mit den a-posteriori prädiktiven Erwartungswerten $\mathbb{E}(\theta^*|y)$. Wie wir in den Boxplots in Abb. B.4 und den Tabellen B.6 und B.7 sehen, gibt es keine Unterschiede zwischen den gewählten a-priori Verteilungen und dem ungetrimmten und getrimmten Modell bezüglich $\mathbb{E}(\theta^*|y)$. Alle Werte sind fast identisch. Die Erwartungswerte sind symmetrisch um die wahren Log Odds Ratios $\mu_0 = -0.693$ bzw. $\mu_0 = -0.357$ verteilt. Daran können wir sehen, dass die historische Datengenerierung wie vorgesehen funktioniert hat. Wie in Gleichung (2.10) dargestellt, stimmen die

a-posteriori prädiktiven Erwartungswerte mit den a-posteriori Erwartungswerten überein. Es handelt sich um Punktschätzer, für die wir als hyper a-priori Verteilung bezüglich μ die sehr schwach-informative Normalverteilung $N(0, 10^6)$ gewählt haben (vgl. Formel (2.41)). Gerade weil diese Verteilung wenig informativ ist und die hyper a-priori Verteilung für τ eine angemessene Streuung zulässt, stimmt $\mathbb{E}(\theta^*|y)$ mit dem beobachteten gepoolten Log Odds Ratio überein (vgl. Formeln (2.8), (2.15)). Da τ auf den Punktschätzer, den a-posteriori prädiktiven Erwartungswert von θ^* , fast keinen Einfluss hat, waren auch keine Unterschiede zwischen den gewählten a-priori Verteilungen für τ bezüglich dieses Parameters zu erwarten. Unsere Beobachtungen decken sich mit den Erkenntnissen in *Lambert et al.* [40, S. 2409], wo ebenfalls keine Unterschiede bei den Punktschätzern auftraten.

Interstudienvarianz τ^2 . Anders verhält sich dies bei der Interstudienvarianz τ^2 . Die Aussagen für τ^2 sind in Setting A und B identisch. In den Abbildungen B.5 und den Tabellen B.8, B.9 sind jeweils die a-posteriori Erwartungswerte der Standardabweichung τ der einzelnen Metaanalyse-Daten dargestellt. Wenn wir im Folgenden von τ sprechen, sind immer die a-posteriori Erwartungswerte $\mathbb{E}(\tau|y)$ gemeint. Anhand der Boxplots in Abbildung B.5 lässt sich bei den einzelnen Verteilungen gut erkennen, wie τ durch Erhöhung der Studienzahl immer mehr durch die Daten geprägt wird. Der Einfluss der a-priori Verteilungen lässt nach. Die am wenigstens informative Verteilung $U(0, 100)$ benötigt hierfür die meisten Beobachtungen. Die wahre Interstudienvariation τ wurde von uns in beiden Szenarien fest auf 0.5 gesetzt. Wir sehen, dass die Boxplots sich diesem Wert, erkennbar an der Referenzlinie, annähern. Hieran lässt sich ebenfalls feststellen, dass die Simulation der historischen Ereigniszahlen im Mittel über alle Metaanalysekonstellationen korrekt gelungen ist. Die Gleichverteilung verhält sich sehr konservativ und schätzt τ regelmäßig zu hoch. Die beobachteten Log Odds Ratios der wenigen Studien genügen noch nicht, um die a-priori Verteilung ausreichend abzuändern und einzuschränken. Die Half-Normal- und die Half-Cauchy-Verteilung zeigen die geringsten Schwankungen. Ab 4 Studien verändert sich τ nur noch unmerklich. Während sich bei der Half-Normal-, Half-Cauchy- und Gleichverteilung der Median und der Mittelwert der a-posteriori Erwartungswerte von τ langsam dem wahren Wert 0.5 nähern, befindet sich der Median bei der Inverse-Gamma-Verteilung bereits ab 4 Studien unterhalb des wahren Wertes. Im Mittel wird τ systematisch zu klein geschätzt. Auffällig ist außerdem, dass obwohl der Großteil der a-posteriori Erwartungswerte niedriger ist als bei den anderen Verteilungen, die Maxima deutlich höher sind. Wenn wir nun noch berücksichtigen, dass dies nur die mittleren Standardabweichungen sind und dass es somit sehr viel extremere Iterationswerte in den

Markovketten für τ geben muss, wird das Ausmaß der Unterschätzung von τ noch klarer. Da τ direkten Einfluss auf V_τ und somit auf ω bzw. N_{eff} und N_{prior} hat, erklärt dies, warum ω und N_{eff} bei der Inverse-Gamma-Verteilung größer sind. Die Unterschiede zum getrimmten Modell sind vernachlässigbar, weil die Markovketten bezüglich der Werte von θ^* getrimmt wurden.

A-posteriori prädiktive Varianz V_τ . Die Schätzungen der Varianz V_τ haben den größten Einfluss auf den Extrapolationswert der historischen Daten. Sie beeinflussen unmittelbar den Gewichtungsfaktor ω und damit auch N_{eff} . Die Varianzen des zukünftigen Studieneffekts verhalten sich ähnlich wie die a-posteriori Erwartungswerte für τ . Allerdings sind die Auswirkungen der Startverteilungen auf V_τ noch drastischer. Anhand der Boxplots in Abbildung B.6 sehen wir, dass bei 3 und 4 Studien sehr große Standardabweichungen entstehen. Die Werte bei der Gleichverteilung passen sich erst ab 6 Studien denen der anderen Verteilungen an. Die Gleichverteilung verhält sich auch für diesen Parameter konservativ. Bei 3 Studien beträgt $V_\tau^{1/2}$ unter Verwendung von $U(0,100)$ in Szenario A im Median 8.01 und in Szenario B im Median 7.86 (Tab. B.10, B.11). Die Präzision geht gegen 0. Die historischen Daten beinhalten ergo keine Information über den Therapieeffekt einer zukünftigen Studie. Somit wird V_τ nahezu immer höher geschätzt, als mit den übrigen Startverteilungen. Die einzige Ausnahme stellt hier die unbeschränkte Inverse-Gamma-Verteilung dar. Mit $\Gamma^{-1}(0.001,0.001)$ werden in beiden Szenarien bei 3 Studien Standardabweichungen $V_\tau^{1/2}$ größer als 10 geschätzt. Diese Werte sind in realen Situationen undenkbar. Die Variation bezieht sich auf die logarithmierten Werte. Der wirkliche Therapieeffekt, das Odds Ratio, könnte damit nach [62, S. 169] irgendwo im verhältnismäßigen „Range“ (vgl. Kap. 5.1) von $\exp(3.92\tau) = \exp(39.2) = 1.43 \cdot 10^{16}$ sein. Anders gesagt, das wahre Odds Ratio kann fast alle möglichen Werte zwischen 0 und ∞ annehmen. In diesem Kontext ist daher auch keine sinnvolle Vorhersage eines zukünftigen Studienergebnisses möglich. Obwohl bei der Inverse-Gamma-Verteilung bei 3 und 4 Studien immer wieder vereinzelt hohe Werte vorkommen, sind die minimalen Werte bereits niedriger als bei den anderen Verteilungen. Dies führt zu einer Überschätzung der Präzision von θ^* . Den historischen Ergebnissen wird somit mehr Bedeutung zugesprochen als sie tatsächlich über die Genauigkeit des zukünftigen Studieneffekts beinhalten. Der Einfluss der α -Trimming-Methode ist bei der Inverse-Gamma- und der Gleichverteilung beim Vorliegen von nur 3 und 4 Studien groß. Durch das Entfernen extremer Iterationsschritte bezüglich θ^* wird konsequenterweise $V_\tau^{1/2}$ kleiner. Je größer $V_\tau^{1/2}$ bereits im vollen Modell geschätzt wird, desto größer ist der Unterschied zur getrimmten Varianz. Da bei $U(0,100)$ und $\Gamma^{-1}(0.001,0.001)$ sehr hohe

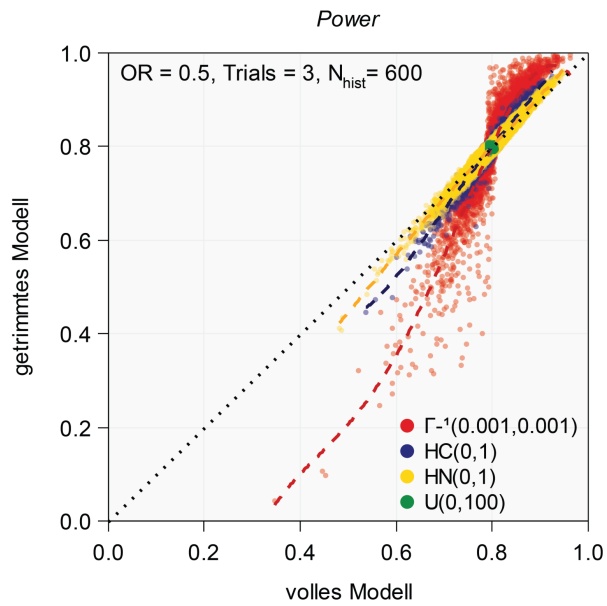
Varianzen bzw. einzelne Ausreißer generiert werden, ist die Auswirkung des Trimmens bei beiden Verteilungen stark. Bei der Half-Normal- und der Half-Cauchy-Verteilung sind die Differenzen zwischen den Modellen geringer. Durch die schwach-informativen $HN(0,1)$ - und $HC(0,1)$ -Verteilungen wird die Varianz bereits im Vorfeld auf einen realistischen Wertebereich eingeschränkt. Daher können keine extremen Ausreißer entstehen. Die Werte für $V_\tau^{1/2}$ bei der Half-Normal-Verteilung streuen insgesamt am wenigstens (Min=0.22, Max=1.79). Ab 4 Studien verändert sich $V_\tau^{1/2}$ nur noch unmerklich. Ähnlich verhalten sich die Werte bei der Half-Cauchy-Verteilung (Min=0.25, Max=5.61). Die Boxen in Abbildung B.6 liegen nahezu auf dem gleichen Niveau wie die der Half-Normal-Verteilung. Allerdings lässt die Half-Cauchy-Verteilung im Zusammenspiel mit den Daten weiter höhere Werte zu. Die Wahrscheinlichkeitsmasse für Werte im Bereich $[1, \infty)$ ist bei der $HC(0,1)$ -Verteilung ein wenig höher als bei der $HN(0,1)$ -Verteilung.

Bayesianische Power. Um die Wirkung der a-posteriori prädiktiven Parameter und der Precision Weight Effective Sample Sizes auf das Ergebnis der neuen Studie zu verdeutlichen, betrachten wir die simulierte Power (vgl. Kap. 5.1.5). Da in Szenario A die benötigte Fallzahl der neuen Studie $N_{ref} = 270$ beträgt, ist N_{eff} im Verhältnis zu N_{ref} groß. Der potentielle Einfluss der Precision Weight Effective Sample Size auf die Power der neuen Studie ist somit sehr hoch. In Szenario B ist $N_{ref} = 1074$ und damit das Verhältnis zwischen virtueller und benötigter Fallzahl geringer. Dementsprechend ist in Szenario B eine geringere Veränderung der Power durch die Reduzierung der Fallzahl um N_{eff} Patienten zu erwarten. Eine sinnvolle Interpretation der Power können wir nur im Zusammenspiel mit den anderen Parametern $\mathbb{E}(\theta^*|y)$, V_τ und N_{eff} vornehmen. Trotzdem geben wir an dieser Stelle zunächst einen deskriptiven Überblick über die Power. In Abbildung B.5 ist die Power unter Verwendung der Ergebnisse aus dem vollen Modell und die Power unter Verwendung der getrimmten Ergebnisse dargestellt. Als Grundlage der „wahren“ Power wurden die Schätzer des vollen Modells verwendet und die Referenzfallzahl um N_{true} reduziert. Wir sehen auf den ersten Blick, dass die Power bei Verwendung aller Verteilungen stark schwankt. Mit Erhöhung der Studienzahl steigt die Variation. In Tabelle B.8 und B.9 sind die Quantile der ermittelten Power dargestellt. Die Färbung von rot über grün zu blau stellt die Einhaltung der Power dar. Die Orange- und Rottöne repräsentieren einen hohen Powerverlust. Die Gleichverteilung zeigt bei 3 Studien keine Veränderung der Power. Es bleibt konstant bei ca. 80%, was daran liegt, dass die Fallzahl nur um ca. 1 Patient reduziert wird. Bei Anwendung der Inverse-Gamma-Verteilung zeigt sich bereits bei niedriger Studienzahl ein sehr großen Rück-

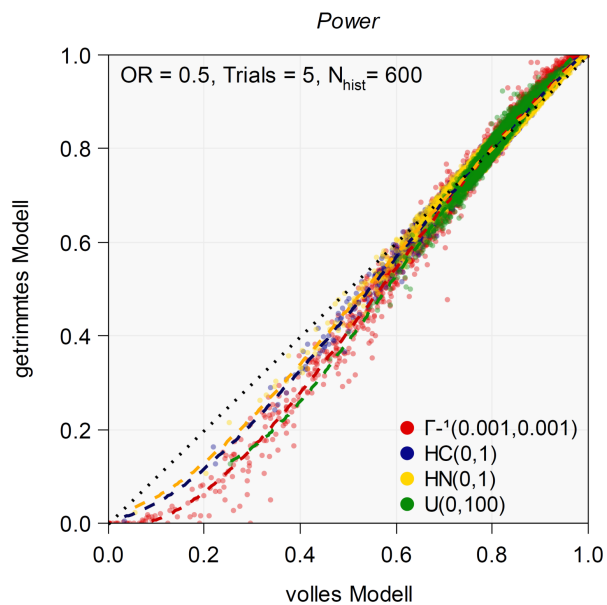
gang in der Power. Das 1% Quantil ist ab 4 Studien durchgängig niedriger als bei den anderen Verteilungen.

Ein Powerverlust oder -gewinn ist logischerweise genau dann zu erwarten, wenn gleichzeitig die Varianz des historischen Studieneffekts klein ist, aber der vorhergesagte zukünftige Studieneffekt weit vom hypothetischen Effekt der neuen Studie abweicht. Wenn der informative a-priori Varianzparameter sehr klein ist, so kann N_{eff} fast so groß wie N_{ref} werden. Wenn aber gleichzeitig der informative a-priori Studieneffekt stark verschoben ist, kann die Power sogar auf 0 absinken. Der Vergleich bei geringer Studienanzahl zwischen der Power mit ungetrimmter und getrimmter Fallzahl, unter Verwendung der jeweils ungetrimmten bzw. getrimmten a-posteriori prädiktiven Parameter als informative Prior, zeigt klare Unterschiede. Die Power unter Verwendung des getrimmten Modells weist deutlich mehr Streuung auf. In den Streupunktdiagrammen B.8, von denen in Abbildung 5.2 ein Auszug wiedergegeben ist, sehen wir, wie das Powerdefizit bzw. der Powerbenefit mit Erhöhung der Studienzahl zunimmt. Wenn die Precision Weight Effective Sample Size genau der a-posteriori prädiktiven Präzision entspräche, so wäre zu erwarten, dass alle Punkte auf einer Geraden liegen, die etwas steiler als die Winkelhalbierende verläuft. Je robuster die Schätzungen, desto mehr sollte die Gerade der Winkelhalbierenden gleichen.

Abbildung 5.2a zeigt, dass bei der Inverse-Gamma-Verteilung durch das Trimmen sehr viel Power verloren geht, falls die ungetrimmte Power bereits niedriger als 80% ist. Andererseits ist die Powerzunahme im getrimmten Modell groß, wenn die ungetrimmte Power größer als 80% ist. Bei den übrigen Verteilungen ist dieser Unterschied zwar ebenfalls vorhanden, jedoch nicht so stark ausgeprägt. Bei der Half-Normal-Verteilung formen die Punkte schon bei 3 Studien eine Gerade, sodass die Korrelation zwischen ungetrimmter und getrimmter Power sehr hoch ist. Die genaueren Zusammenhänge und Begründungen erläutern wir im nächsten Abschnitt. In Abbildung 5.2b wird ersichtlich, dass die Abweichungen zwischen getrimmtem und ungetrimmtem Modell bei 5 Studien nur noch marginal sind. Die Regressionskurven, polynomiale Regressionen vom Grad 5, nähern sich ab 5 Studien immer mehr einer Geraden und der Winkelhalbierenden an. Größere Powerabweichungen zwischen den beiden Modellen beobachten wir nur noch, wenn die ungetrimmte Power niedriger als 50% ist. Die Power im getrimmten Modell ist dann nochmals deutlich niedriger. Dieser Effekt ist allerdings vernachlässigbar, da wir bei der tatsächlichen Durchführung einer neuen klinischen Studie bereits die Parameterschätzungen aus der meta-analytisch prädiktiven Methode kennen. Folglich würden wir nicht wie hier verfahren, sondern die neue Studie basierend auf dem geschätzten



(a) Vergleich bei 3 Studien .



(b) Vergleich bei 5 Studien .

Abb. 5.2: Scatterplots: Vergleich der bayesianische Power im vollen und im getrimmten Modell in Szenario A bei 3 und 5 Studien.

Behandlungseffekt der historischen Studien planen. Die Ergebnisse für diese Vorgehensweise stellen wir in Abschnitt 5.3.2 dar.

Multivariable Zusammenhänge Die Ergebnisse der Einzelparameter lassen sich erst in einem multivariablen Zusammenhang richtig interpretieren. Wann kann der Gewichtungsfaktor ω bzw. die Fallzahl N_{eff} groß werden? Genau dann, wenn die Interstudienvarianz τ^2 klein ist und folglich die a-posteriori prädiktive Varianz V_τ klein ist. In diesem Fall besitzt θ^* eine hohe Präzision und der Informationsgehalt der historischen Studien erhält einen großen Wert. Wenn nur wenige historische Studien vorliegen, reichen die Daten nicht aus, um eine zutreffende Vorhersage über den Parameter der neuen Studie zuzulassen. Die historischen Ergebnisse werden somit stark inflationiert. Konsequenterweise ist N_{eff} niedrig. Insbesondere trifft dies zu, wenn wir sehr schwach-informative a-priori Verteilungen für τ wie die Gleichverteilung $U(0,100)$ benutzen. Die a-posteriori prädiktive Verteilung wird dann nicht durch die Likelihood bestimmt, sondern nahezu allein durch die a-priori Verteilung. Bei 3 und 4 historischen Studien in Szenario A und B beträgt N_{eff} unter Anwendung der Gleichverteilung im Mittel bei 0 bis 7 Patienten. Die Gleichverteilung verhält sich sehr konservativ.

In Abbildung B.9 erkennen wir den Zusammenhang zwischen der Power und dem vorhergesagten Behandlungseffekt θ^* auf Grundlage der historische Studien. In Abbildung 5.3 sind auszugsweise die Ergebnisse für 6 Studien für das Szenario A dargestellt. Die Power steigt, wenn der a-posteriori prädiktive Erwartungswert kleiner als das geplante Log Odds Ratio ist. Der Therapieunterschied ist dann größer als angenommen. Ist θ^* andererseits größer als der hypothetische Behandlungseffekt $\mu_0 = -0.693$, also näher bei 0, fällt die Power exponentiell

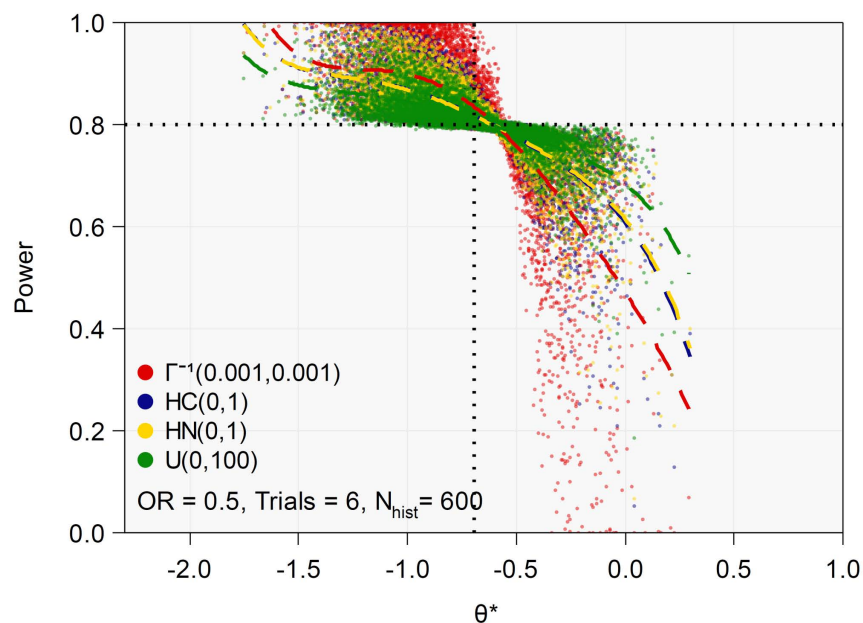


Abb. 5.3: Scatterplot: Vergleich der a-posteriori prädiktiven Erwartungswerte von θ^* und der bayesianischen Power in Szenario A bei 6 Studien.

ab. Die Regressionskurven sind für Polynome vom Grad 5 angepasst. Die Half-Normal- und Half-Cauchy-Verteilung verhalten sich identisch. Dies liegt daran, dass die Varianz V_τ und damit die virtuelle Fallzahl N_{eff} bei beiden Verteilungen in etwa gleich sind. Die ausgeprägteste Powerreduktion tritt bei der Inverse-Gamma-Verteilung auf. Hier wurde V_τ im Durchschnitt am kleinsten geschätzt. Falls der gepoolte Therapieeffekt kleiner als angenommen oder sogar gegenläufig ist und gleichzeitig alle historischen Studien sehr homogen sind, also die Interstudienvarianz klein ist, wird θ^* zwar sehr präzise, aber auch hinsichtlich des wahren Therapieeffekts sehr verfälscht geschätzt. Das Resultat ist, dass die historische Information stark gewichtet wird. Dadurch ist ω bzw. N_{eff} groß, der vorhergesagte Behandlungsunterschied allerdings präzise falsch. Wird dieser verschobene Effekt mit einer hohen Präzision in die neue Studie eingebracht, müssen durch die Reduzierung der Fallzahl N_{ref} um die hohe Precision Effective Sample Size N_{eff} nur noch wenige neue Patienten rekrutiert werden. Die Beobachtungen der neuen Patienten haben keine Chance den verschobenen stark-informativen a-priori Erwartungswert zu korrigieren. Die Konsequenz ist ein drastischer Powerverlust oder -gewinn (siehe Abb. B.10). Dieser tritt bei allen Verteilungen auf. Das Ausmaß des Powergewinns bzw. Powerverlusts hängt stark von der Anzahl N_{eff} ab. Da bei der Inverse-Gamma-Verteilung die Varianz tendenziell kleiner und deshalb ω und N_{eff} höher geschätzt werden, tritt dieser Effekt bei ihr am massivsten auf. Die Regressionskurve (rot) verläuft steiler und die Punkte streuen sehr weit (vgl. Abb. 5.3). Bei der Gleichverteilung ist die Präzision und damit der Gewichtungsfaktor ω und die Fallzahl N_{eff} am kleinsten. Darum beobachten wir hier den kleinsten Zugewinn bzw. Verlust an Power.

Bei wenigen Studien sind die Auswirkungen auf die Power niedrig. Die Varianz von θ^* ist diesbezüglich größer und folglich N_{eff} kleiner. Die neue Studie wird um weniger Patienten reduziert, sodass dementsprechend mehr neue Patienten rekrutiert werden. Die Outcomes der neuen Patienten haben gegenüber der verzerrten a-priori Information ein höheres Gewicht. Damit ist der Verlust bzw. Gewinn an Power bei wenigen Studien geringer. Die Regressionskurven in Abbildung B.9 bei 3 Studien verlaufen darum flacher und weichen weniger von der Referenzlinie 80% ab. Weil bei der Gleichverteilung die Daten der 3 historischen Studien fast kein Gewicht haben und somit kaum weniger Patienten (max. $N_{eff} = 1.6$) rekrutiert werden, verläuft die Regressionkurve (grün) genau auf der Referenzlinie. In Szenario B tritt der beschriebene Effekt ebenfalls auf, fällt jedoch im Vergleich zu Szenario A geringer aus. Dies liegt daran, dass die benötigte Fallzahl der neuen Studie ohne a-priori Information $N_{ref} = 1074$ verhältnismäßig groß zur vorhergesagten, reduzierbaren Patientenzahl N_{eff} ist. Der geplante Behandlungseffekt $OR = 0.7$ liegt näher an 1, sodass für den Nachweis der einseitigen

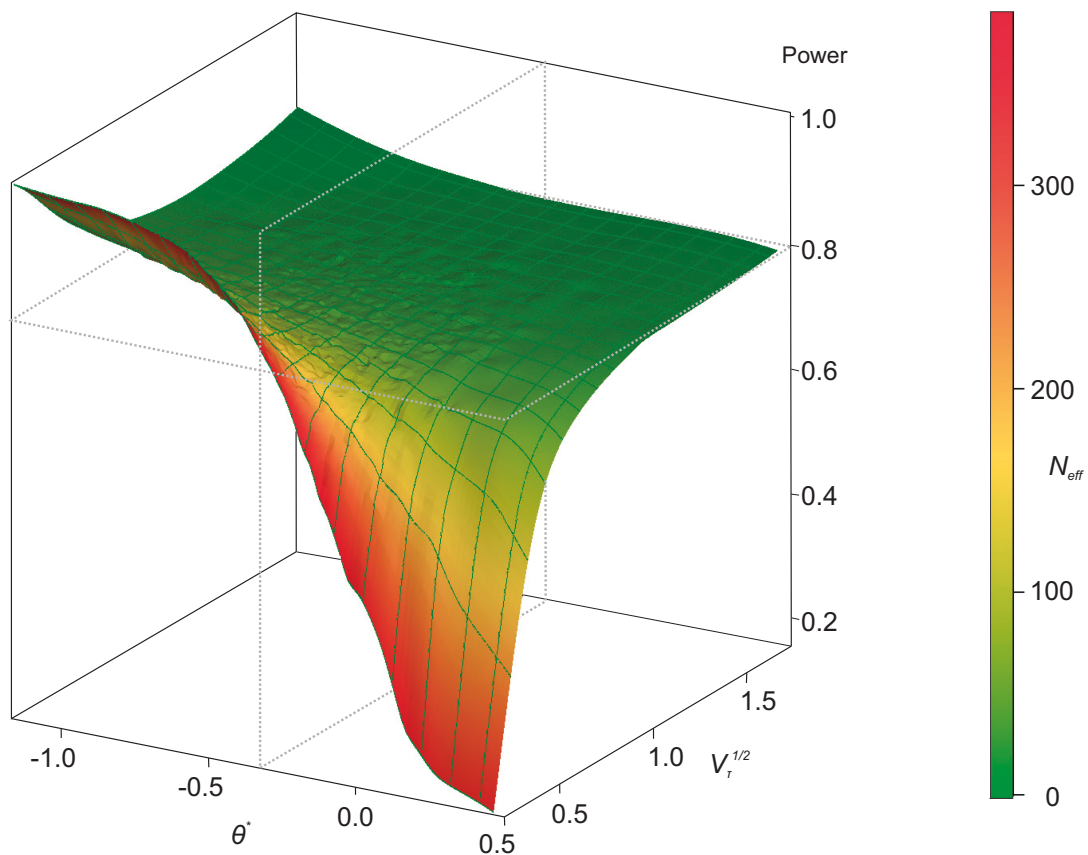


Abb. 5.4: Zusammenhang zwischen $\mathbb{E}(\theta^*|y)$, $V_\tau^{1/2}$, N_{eff} und bayesianischer Power.

Szenario A, $OR = 0.5$, $Trials = 6$, $\tau^2 \sim \Gamma^{-1}(0.001, 0.001)$. A-posteriori prädiktiver Mittelwert θ^* und Standardabweichung $V_\tau^{1/2}$, Precision Weight Effective Sample Size des vollen Modells N_{eff} und simulierte bayesianische Power. Die gestrichelten Ebenen markieren die geplante Power von 80% und den angenommenen Behandlungseffekt $\mu_0 = -0.693$.

Hypothese (5.1) ein kleineres 97.5%-Konfidenzintervall als in Szenario A benötigt wird. Die eingebrachte a-priori Varianz ist für die Hypothese in Szenario B vergleichsweise groß, sodass die Reduzierung um N_{eff} Patienten in Relation zu N_{ref} niedrig ist. Die Power kann folglich nicht so tief absinken.

Exemplarisch sind die Abhängigkeiten zwischen den a-posteriori prädiktiven Parametern $\mathbb{E}(\theta^*|y)$ und V_τ sowie der virtuellen Fallzahl N_{eff} und der Power der neuen Studie in Abbildung 5.4 veranschaulicht. Der vierdimensionale Zusammenhang ist hier am Beispiel der Inverse-Gamma-Verteilung für das Szenario A mit 6 historischen Studien abgebildet. Die Powerveränderungen sind immer dann groß, wenn $\mathbb{E}(\theta^*|y)$ vom geplanten Effekt abweicht, gleichzeitig die Standardabweichung $V_\tau^{1/2}$ klein ist und dementsprechend die virtuelle Patientenzahl N_{eff} hoch ist. Diese Darstellungsform stellt alle relevanten Zusammenhänge der Parameter am anschaulichsten dar.

Prior Effective Sample Size N_{prior} . Die Prior Effective Sample Size N_{prior} wird unmittelbar von den a-priori Parametern beeinflusst. Darum zeigen sich hier die gleichen Unterschiede wie bei N_{eff} bezüglich der verwendeten hyper a-priori Verteilungen für den Varianzparameter im meta-analytisch prädiktiven Ansatz. Ebenso ist dies bei den Abweichungen von N_{prior} bei Anwendung der ungetrimmten Parameter und N_{prior_trim} bei Anwendung der getrimmten Parameter. Wie wir in Abbildung B.11 erkennen, schwankt N_{prior} beträchtlich.

Da wir in diesem Setting so wie in Bemerkung 5.1 verfahren und die historischen Vorhersagen bei der Planung nicht berücksichtigen, hat dies entscheidende Auswirkungen auf N_{prior} . Falls wir beispielsweise einen viel kleineren oder sogar der Annahme gegenläufigen informativen a-priori Behandlungseffekt in die Auswertung der neuen Studie einbringen, kann die benötigte Fallzahl für eine nominelle Power von 80% größer als N_{ref} werden. Das heißt, wir können keine Patienten einsparen, sondern müssen noch zusätzliche Teilnehmer rekrutieren. Nach Definition 3.1 ist N_{prior} dann negativ. Dieser Fall tritt in allen betrachteten Simulationsszenarien ein. Die deskriptiven Maßzahlen sind in den Tabellen B.14 und B.15 aufgelistet. In der extremsten Situation müssen zusätzlich zur ursprünglichen Fallzahl N_{ref} noch einmal fast genauso viele Patienten rekrutiert werden (Szenario A, 5 Studien, $\Gamma^{-1}(0.001, 0.001)$: $N_{prior} = -262$). In fast allen Szenarien, außer bei 3 und 4 Studien mit $U(0, 100)$, wird die Fallzahleinsparung N_{prior} so groß wie die benötigte Fallzahl N_{ref} . Infolgedessen ist die a-priori Präzision ausreichend groß bzw. der a-priori Erwartungswert so viel kleiner als der angenommene Behandlungseffekt, dass keine neuen Patienten rekrutiert werden müssen. Bei der Gleichverteilung ist die a-priori Varianz noch zu groß, sodass bei 3 Studien in Szenario A maximal 3 Patienten eingespart werden können bzw. maximal 4 zusätzliche Patienten rekrutiert werden müssen.

Wir wissen bereits, dass die a-posteriori prädiktiven Erwartungswerte $\mathbb{E}(\theta^*|y)$ der Metaanalysen im Median und Mittel mit den wahren Log Odds Ratios übereinstimmen und zumindest bei der Half-Normal- und Half-Cauchy-Verteilung auch die a-posteriori Erwartungswerte $\mathbb{E}(\tau|y)$ im Median die wahre Interstudienstandardabweichung richtig schätzen. Für diese beiden Verteilungen sollte der Median von N_{prior} ungefähr die richtige Fallzahleinsparung repräsentieren. Tendenziell können wir im Mittel über alle Metaanalysekonstellationen vermehrt Patienten einsparen. Mit der Half-Normal-Verteilung ist N_{prior} in Szenario A im Median 28, 35, 40, 46 Patienten und in Szenario B im Median 30, 34, 40, 48. Die Methode stellt daher im Mittel einen Gewinn dar, auch wenn sie der neuen Studie im Sinne von zusätzlichen Patienten schaden kann. Auf der anderen Seite kann sie aber auch vermehrt dazu führen, dass keine Neurekrutierung von Patienten notwendig ist. In der realen Situation können wir das Pro-

blem einer zusätzlichen Patientenrekrutierung umgehen, indem wir zunächst die historischen Daten betrachten und danach die neue Studie genau bezüglich des a-posteriori prädiktiven Erwartungswertes planen.

In Abbildung B.11a können wir beispielsweise in Szenario A bei 3 Studien einen Bereich zwischen ca. 200 und 276 Patienten erkennen, in den keine Werte von N_{prior} fallen. Diese ausgelassenen Wertebereiche treten bei allen Studienanzahlen und ebenso in Szenario B auf. Dieser Bereich repräsentiert die Situationen, in denen die Durchführung der neuen Studie keinen Sinn macht. Eine Studie, in der weniger als beispielsweise 50 neue Patienten rekrutiert und dazu sehr informative a-priori Parameter verwendet werden, führt dazu, dass ein bereits ohne Neurekrutierung signifikantes Ergebnis durch die wenigen neuen Patienten nicht mehr signifikant wäre. In dieser Konstellation ist die Power aufgrund eines verschobenen informativen a-priori Erwartungswert oder einer großen a-priori Präzision bereits ohne neue Patienten größer als 80%. Wenn nun wenige zusätzliche Patienten rekrutiert werden, beeinflusst der Standardfehler des angenommenen Log Odds Ratios die Power (vgl. Formeln (3.27), (3.29)). Der Standardfehler der neuen Studie wird mit Gleichung 3.12 berechnet und ist aufgrund der kleinen Fallzahl sehr groß. Die a-posteriori Präzision (vgl. Formel 2.7) besteht aus der Summe der a-priori Präzision und der Präzision der Daten und geht in die Powerformel ein. Ebenso ist der a-posteriori Erwartungswert ein durch die Präzisionen gewichtetes Mittel des geplanten Log Odds Ratios der neuen Studie und des a-priori Erwartungswerts. Durch den großen Standardfehler kann einerseits die a-posteriori Varianz durch die Neurekrutierung wieder vergrößert werden oder andererseits der a-posteriori Erwartungswert wieder näher an das angenommene Log Odds Ratio verschoben werden. In beiden Fällen wird der Ausdruck $\mu_1 \tau_1^{-1}$ in Gleichung (3.26) und Gleichung (3.27) zunächst größer. Dadurch wird es also schwieriger ein signifikantes Ergebnis zu erhalten. Folglich wird die Power vorerst kleiner, sodass sie 80% unterschreiten kann. Erst wenn ausreichend viele neue Patienten eingeschlossen werden, wird der beobachtete Standardfehler und damit die a-posteriori Varianz wieder klein genug, damit die Power das Niveau von 80% überschreitet. In Szenario B ist gibt es ebenfalls bei allen Studienanzahlen einen Bereich, der nicht von N_{prior} angenommen wird. Daraus resultiert die insgesamt linksschiefe Verteilung von N_{prior} .

5.3.2. Ergebnisse bei adjustiertem a-priori Erwartungswert zur Validierung von N_{eff}

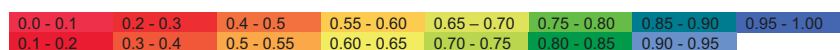
In dieser zusätzlichen Analyse adjustieren wir den a-priori Erwartungswert des Normal-Normal-Modells der neuen Studie auf den jeweils angenommenen Effekt (vgl. Kap. 5.1.5). Wir wollen hiermit überprüfen, ob der Wert der a-posteriori prädiktiven Varianz, die als a-priori Varianz in die neue Studie eingeht, gleichbedeutend mit N_{eff} ist. In dieser ergänzenden Poweranalyse, sollte die a-priori Präzision demnach exakt die Anzahl der weniger rekrutierten Patienten N_{eff} ausgleichen, sodass die Power trotz reduzierter Fallzahl 80% beträgt. Die intuitive Umrechnung des Gewichtungsfaktors der Präzisionen in die Precision Weight Effective Sample Size N_{eff} wäre dann korrekt.

In Tabelle 5.2 ist die simulierte Power deskriptiv dargestellt. Im vollen Modell mit der Gleichverteilung wurden vergleichsweise große Varianzen V_τ und niedrige Fallzahlen N_{eff} geschätzt. Im Gegensatz dazu wurden mit der Inverse-Gamma-Verteilung kleine Varianzen und hohe

		Min	1% Quantil	5% Quantil	10% Quantil	25% Quantil	Median	75% Quantil	90% Quantil	95% Quantil	99% Quantil	Max
<i>Prior distribution for τ</i>												
$\Gamma^{-1}(0.001, 0.001)$	Power with $\mu_0 = -0.693$	0.785	0.790	0.793	0.795	0.801	0.810	0.831	0.867	0.894	0.946	1.000
	Power _{trim} with $\mu_0 = -0.693$	0.786	0.790	0.794	0.797	0.804	0.816	0.844	0.888	0.921	0.982	1.000
HC(0,1)	Power with $\mu_0 = -0.693$	0.784	0.791	0.794	0.796	0.801	0.808	0.818	0.832	0.843	0.866	0.916
	Power _{trim} with $\mu_0 = -0.693$	0.786	0.792	0.795	0.797	0.802	0.810	0.821	0.838	0.850	0.877	0.932
HN(0,1)	Power with $\mu_0 = -0.693$	0.788	0.793	0.796	0.798	0.803	0.810	0.820	0.832	0.842	0.863	0.908
	Power _{trim} with $\mu_0 = -0.693$	0.787	0.793	0.796	0.798	0.803	0.811	0.822	0.836	0.847	0.872	0.922
U(0,100)	Power with $\mu_0 = -0.693$	0.785	0.789	0.791	0.793	0.797	0.802	0.807	0.815	0.823	0.842	0.883
	Power _{trim} with $\mu_0 = -0.693$	0.785	0.790	0.792	0.793	0.796	0.803	0.809	0.820	0.830	0.852	0.909

(a) Szenario A: $OR = 0.5$, mit festem a-priori Parameter $\mu_0 \equiv -0.693$ für die Poweranalyse .

		Min	1% Quantil	5% Quantil	10% Quantil	25% Quantil	Median	75% Quantil	90% Quantil	95% Quantil	99% Quantil	Max
<i>Prior distribution for τ</i>												
$\Gamma^{-1}(0.001, 0.001)$	Power with $\mu_0 = -0.357$	0.793	0.796	0.797	0.798	0.800	0.802	0.807	0.816	0.823	0.835	0.855
	Power _{trim} with $\mu_0 = -0.357$	0.794	0.796	0.798	0.799	0.800	0.803	0.810	0.820	0.828	0.841	0.863
HC(0,1)	Power with $\mu_0 = -0.357$	0.793	0.796	0.798	0.798	0.800	0.802	0.804	0.808	0.812	0.819	0.836
	Power _{trim} with $\mu_0 = -0.357$	0.794	0.797	0.798	0.799	0.800	0.802	0.805	0.810	0.814	0.822	0.839
HN(0,1)	Power with $\mu_0 = -0.357$	0.794	0.797	0.798	0.799	0.801	0.802	0.805	0.809	0.812	0.818	0.833
	Power _{trim} with $\mu_0 = -0.357$	0.794	0.797	0.798	0.799	0.801	0.803	0.805	0.810	0.813	0.820	0.836
U(0,100)	Power with $\mu_0 = -0.357$	0.794	0.796	0.797	0.798	0.799	0.800	0.802	0.804	0.807	0.813	0.826
	Power _{trim} with $\mu_0 = -0.357$	0.793	0.796	0.797	0.798	0.799	0.800	0.802	0.806	0.809	0.817	0.832



(b) Szenario B: $OR = 0.7$, mit festem a-priori Parameter $\mu_0 \equiv -0.357$ für die Poweranalyse .

Tab. 5.2: Deskriptive Statistiken für die Power mit fixiertem μ_0 in der neuen Studie.

Für die Poweranalyse wurde die a-posteriori prädiktive Varianz des jeweiligen hierarchischen Modells als informativer a-priori Parameter für das Normal-Normal-Modell der neuen Studie benutzt. Der informative a-priori Erwartungswert ist der angenommene Studieneffekt. Die geplante Fallzahl N_{ref} der neuen Studie wurde um N_{eff} (getrimmt, ungetrimmt) reduziert.

Fallzahlen generiert. Zur Beantwortung unsere Frage können wir daher die Ergebnisse für alle Verteilungen sowie für alle getrimmten und ungetrimmten Parameter zusammenfassen. Auf die Aussage zur Gleichwertigkeit von N_{eff} und V_τ hat die Wahl der hyper a-priori Verteilung des vollen Modells keinen Einfluss.

In Tabelle 5.2 erkennen wir, dass die Power höchstens um 1.6% sinkt. Die minimale Power beträgt 78.4% und die maximale Power 100%. In diesen Fällen hätte N_{eff} sogar etwas größer ausfallen können. Der Verlust an Power ist durchaus tolerierbar. Besonders wenn N_{eff} klein ist, sind Schwankungen im Umfang von ca. 1% im Rahmen von Rundungsfehlern bei dieser Simulation zu erwarten. Die Fallzahl wird immer auf ganze Patienten reduziert. Entsprechend kann die Power geringfügig absinken oder zunehmen. Insgesamt liegt die Power in beiden Szenarien A und B wortwörtlich im grünen Bereich. Tendenziell ist vermehrt ein Powergewinn zu betrachten. Das bedeutet, dass die Umrechnung von V_τ zu N_{eff} eher konservativ ausfällt. Die Precision Weight Effective Sample Size könnte in einigen Fällen sogar etwas höher kalkuliert werden. Allerdings ist die etwas konservative Umrechnung gut, da N_{eff} unabhängig vom Design der neuen Studie bestimmt wird. Der tatsächliche durch die informativen a-priori Parameter übertragene Informationsgehalt ist aber immer von den Planungsannahmen der neuen Studie abhängig. Für unsere Szenarien können wir aber die Schlussfolgerung treffen, dass die Übersetzung des Informationsgehaltes in die virtuelle Fallzahl sehr gut, wenn auch eher konservativ funktioniert.

In der zweiten Analyse mit adjustiertem a-priori Erwartungswert haben wir zudem die Prior Effective Sample Size N_{prior} für die speziell geplanten Studien berechnet. Sie repräsentiert die wirkliche Anzahl Patienten, um die die neue Studie unter Verwendung der informativen a-priori Präzision aus dem vollen Modell reduziert werden kann, sodass das Powerniveau von 80% eingehalten wird. Da der a-priori Erwartungswert der angenommen Behandlungseffekt ist, kann die Fallzahl $N_{ref,inf}$, die wir bei Anwendung der informativen Parameter benötigen, die Fallzahl N_{ref} , die wir bei uninformativer Planung brauchen, nicht übersteigen. Daher kann die Prior Effective Sample Size nicht negativ werden und es folgt $N_{prior} \in [0, N_{ref}]$. In Abbildung B.12 und den Tabellen B.16, B.17 sind die Einsparungen getrennt für alle Studienanzahlen und Startverteilungen dargestellt. Aus den bereits beschriebenen Gründen unterscheiden sich die Fallzahlen zwischen den gewählten a-priori Verteilungen und den Studienanzahlen. Diesbezüglich verhält sich N_{prior} genauso wie N_{eff} . Wir können aber jetzt die beiden Fallzahlen direkt gegenüberstellen, indem wir jeweils Tabelle B.1 mit Tabelle B.16 und Tabelle B.2 mit Tabelle B.17 vergleichen. Hier wird ersichtlich, dass N_{prior} fast immer

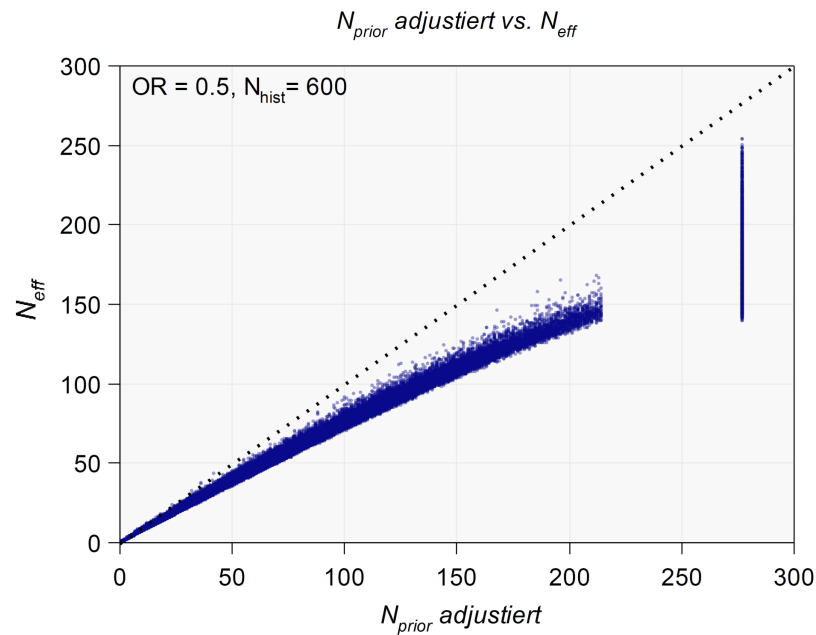
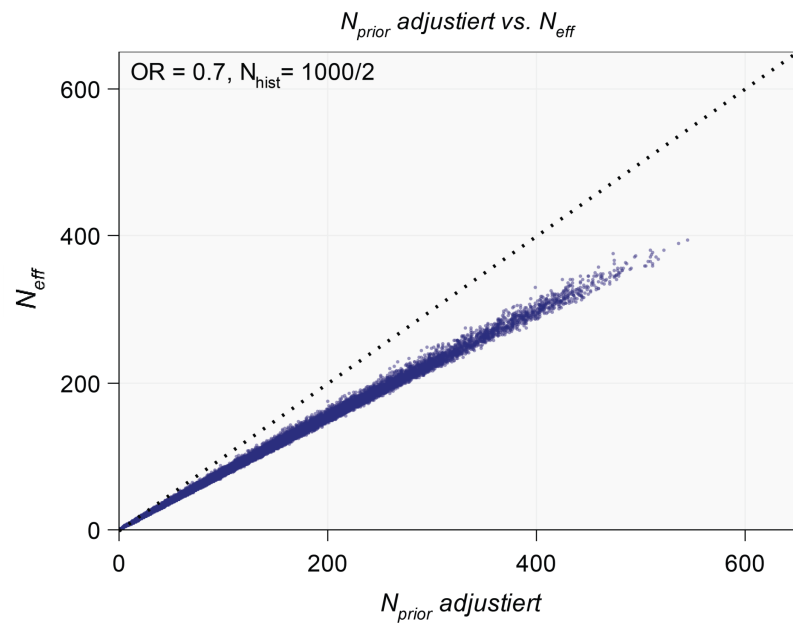
(a) Szenario A: $OR = 0.5$.(b) Szenario B: $OR = 0.7$.

Abb. 5.5: Scatterplots: Zusammenhang zwischen N_{prior} adjustiert und N_{eff} .

Für die Berechnung von N_{prior} wurde die a-posteriori prädiktive Varianz des vollen Modells als informativer a-priori Parameter für das Normal-Normal-Modell der neuen Studie benutzt. Der informative a-priori Erwartungswert ist der angenommene Studieneffekt $\mu_0 = -0.693$ bzw. $\mu_0 = -0.357$.

etwas größer ist als N_{eff} . Die beiden Fallzahlen sind in den Streupunktdiagrammen in Abbildung 5.5 gegeneinander abgetragen. Im Anhang sind die Scatterplots ergänzend getrennt nach Studien und Startverteilungen abgebildet (vgl. Abb. B.13). In den Diagrammen wird noch einmal direkt ersichtlich, dass die beiden Fallzahlen N_{eff} und N_{prior} nicht gleich sind.

Es gibt aber offensichtlich einen linearen Zusammenhang. In beiden betrachteten Szenarien ist N_{prior} fast immer größer als N_{eff} . Dies bestätigt die eher konservative Schätzung von N_{eff} . Tatsächlich beinhaltet also die informative a-priori Präzision mehr Aussagekraft über den zukünftigen Behandlungseffekt als durch N_{eff} reflektiert wird. In Szenario A sehen wir, dass N_{prior} ab einer bestimmten a-priori Präzision gleich der Referenzfallzahl N_{ref} ist. Die a-priori Präzision ist damit so genau, dass das einseitige Konfidenzintervall des Log Odds Ratios bereits ohne zusätzliche Rekrutierung neuer Patienten mit einer Wahrscheinlichkeit von 80% kleiner als 0 ist. Damit muss die neue Studie nicht mehr durchgeführt werden. In Szenario B ist N_{ref} zu groß, um von N_{prior} erreicht zu werden. Die geschätzten a-posteriori prädiktiven Varianzen sind dazu nicht klein genug. In Szenario A sehen wir in Abbildung 5.5a und B.13a einen Bereich zwischen ca. 225 bis 276, der nicht durch N_{prior} angenommen wird. Das liegt an dem bereits diskutierten Phänomen, dass die a-priori Präzision bereits zu informativ ist. Die Durchführung der neuen Studie ist dann nicht mehr nötig. Eine zusätzliche Rekrutierung weniger Patienten würde das bereits signifikante Ergebnis wieder aufheben.

5.3.3. Resümee der Simulationsstudie

Zusammenfassend können wir festhalten, dass wir unseren Simulationsszenarien robuste Parameterschätzungen erhalten, wenn mindestens 5 Studien vorhanden sind. Bei weniger Studien hängen die a-posteriori Parameter noch zu stark von den gewählten hyper a-priori Verteilungen des Varianzparameters ab. Für die Anwendung geeigneter a-priori Verteilungen stellt sich die Frage, ob eine Überschätzung oder eine Unterschätzung des Informationsgehalts der historischen Daten schlimmer ist. Der konservative Ansatz, eine Unterschätzung der a-posteriori prädiktiven Varianz, führt dazu, dass weniger Patienten in einer neuen Studie einspart werden können. Ein konservativer Ansatz ist ethisch gesehen zu befürworten und gegenüber den Gesundheitsbehörden leichter zu rechtfertigen. Falls umgekehrt zu wenige neue Patienten rekrutiert werden, könnten weitere Aussagen zu eventuellen Safety-Kriterien oder zu ergänzenden explorativen Fragestellungen nicht mehr getroffen werden. Die Inverse-Gamma-Verteilung $\Gamma^{-1}(0.001, 0.001)$ unterschätzt die Interstudienvarianz systematisch und eignet sich darum am wenigsten für die praktische Anwendung. Als geeignete Startverteilungen bieten sich hier die $HN(0, 1)$ - und die $HC(0, 1)$ -Verteilungen an.

Insgesamt wird die Power eingehalten, wenn der neu geplante Studieneffekt auf dem a-posteriori prädiktiven Effekt aus den historischen Ergebnissen basiert. Die intuitive Precision Weight Effective Sample Size N_{eff} stimmt nicht exakt mit der Prior Effective Sample Size

N_{prior} überein. Es besteht aber ein linearer Zusammenhang. In den simulierten Szenarien ist N_{eff} kleiner als N_{prior} , was zu einer konservativen Einschätzung der historischen Information durch N_{eff} führt. Im Großen und Ganzen ist die meta-analytisch prädiktive Methode geeignet, um den Informationsgehalt der historischen Studien objektiv zu bewerten. Falls wir eine bestimmte neue Studie zu Grunde legen und N_{prior} verwenden, funktioniert die Übertragung in eine korrespondierende Anzahl virtueller Patienten ebenfalls. Die studienunabhängige Fallzahl N_{eff} eignet sich nur als ein intuitive Maß für die ungefähre Fallzahlersparnis in einer neuen Studie. Dadurch, dass in unserem Kontext N_{eff} vergleichsweise zu niedrig kalkuliert wird, werden in der neuen Studie bei Reduzierung der Fallzahl um N_{eff} eher zu viele als zu wenige neue Patienten rekrutiert. Um die wirkliche Anzahl einzusparender Patienten zu bestimmen, sollte in jedem Fall die Prior Effective Sample Size N_{prior} für die explizite zukünftige Studie bestimmt werden. Die Vorgehensweise in Bemerkung 5.1 ist nicht zu empfehlen, da dann der meta-analytisch prädiktive Ansatz nicht zwingend zu einer Fallzahlreduzierung führt.

IV Diskussion

6 Zusammenfassung und Schlussfolgerungen

Abschließend diskutieren wir noch einmal genauer die Fragestellungen aus der Einleitung. Die Frage, ob die Precision Weight Effective Sample Size, gleichbedeutend mit dem Informationsgehalt der historischen Studien ist, der durch die a-posteriori prädiktive Verteilung in die neue Studie übertragen wird, lässt sich mit näherungsweise beantworten. Trotz der asymptotischen Argumentation und heuristischen Herleitung der Fallzahlformel für N_{eff} , funktioniert die Umrechnung des Informationsgehalts in eine Anzahl virtueller Individuen relativ gut. Eine deutliche Einschränkung ist allerdings, dass N_{eff} unabhängig von den Planungsannahmen einer neuen Studie bestimmt wird. Sie bleibt für jedes Design bei gleicher Zielgröße und Fragestellung gleich. Wir haben aber festgestellt, dass die Fallzahl einer neuen Studie um N_{eff} Patienten reduziert werden, ohne dass die bayesianische Power das geplante Niveau wesentlich unterschreitet. Voraussetzung dafür ist jedoch, dass die neue Studie auf Basis der a-posteriori prädiktiven Parameter geplant wird und diese anschließend als informatives a-priori Wissen in die Auswertung der Studie einfließen (vgl. Kap. 5.3.2).

Um einen Powerverlust unter Verwendung der historischen Information zu vermeiden, muss die neue Studie genau für den Nachweis des geschätzten Behandlungseffekts der historischen Studien geplant werden. Anderenfalls kann das Powerlevel unter Berücksichtigung der früheren Ergebnisse nicht eingehalten werden (vgl. Tab. 4.2, Tab. 4.6, Kap. 5.3.1). Denn ist der nachzuweisende Therapieunterschied beispielsweise kleiner als in den historischen Studien, so ist mit einer Verringerung der Power zu rechnen. Auf diesen Aspekt weist die *FDA* ebenfalls explizit hin:

„If results are very similar, the current study can borrow considerable strength. As current results vary from the previous information, the current study borrows less and less. Very different results borrow no strength at all, or even potentially “borrow negatively”.“
[67, S. 25].

Insofern ist die Extrapolation der Ergebnisse von beiden Behandlungsgruppen auf eine neue klinische Studie limitiert. Allerdings können wir auch nur die Daten der Kontrollgruppe verwenden, so wie es in *Neuenschwander et al.* [50] gemacht wird. Aber genauso müssen wir auch diesbezüglich die Studienplanung am beobachteten Effekt in den historischen Kontrollen ausrichten.

Die nächste Frage, nach der Übereinstimmung von N_{eff} und N_{prior} können wir verneinen. Während sich die Prior Effective Sample Size abhängig von den Studienannahmen wie Signifikanzniveau und Power ändert, bleibt N_{eff} konstant. In den betrachteten Simulationsszenarien ist N_{eff} kleiner als N_{prior} . Somit können in der neuen Studie unter Anwendung der informativen a-priori Parameter mehr Patienten eingespart werden als durch N_{eff} prognostiziert werden. Allerdings ist diese Aussage mit Vorsicht zu genießen, da wir nur spezielle Szenarien betrachtet haben und nicht im Allgemeinen zeigen konnten, dass N_{eff} immer niedriger als N_{prior} und damit konservativ ist. Wir empfehlen daher als Maßzahl für den Extrapolationsgehalt den Gewichtungsfaktor ω oder die Prior Effective Sample Size N_{prior} heranzuziehen und auf die Anwendung von N_{eff} in der neuen Studie zu verzichten.

Die Robustheit des vorhergesagten Behandlungseffekts und damit auch die von N_{eff} und N_{prior} , wird durch die Anzahl der historischen Studien und durch die verwendeten hyper a-priori Verteilungen beeinflusst. Die Frage nach der geeigneten a-priori Verteilung für die Varianzparameter stellt in der Bayes-Community ein viel diskutiertes Problem dar. Eine große Auswahl unterschiedlicher Verteilungen werden von *Lambert et al.* [40] untersucht. Dort wird festgestellt, dass es keine Startverteilung gibt, die in allen Szenarien am besten performt. Dies deckt sich mit unseren Erkenntnissen. Es gibt damit nicht *die* geeignete a-priori Verteilung. Aber es gibt einige, deren Wahl günstiger ist als die anderer. Sollten ausreichend historische Studien vorhanden sein, empfiehlt es sich eine Gleichverteilung als hyper a-priori Verteilung für die Varianzparameter zu verwenden. Wenn aber nur wenige Beobachtungen vorliegen, ist die Gleichverteilung zu vage. Im Rahmen einer Sensitivitätsanalyse sollten für jede Modelanpassung immer mehrere unterschiedliche hyper a-priori Verteilungen betrachtet werden. Diese sollten stets an die entsprechenden Rahmenbedingungen und Zielgrößen angepasst sein. Falls genügend Studien vorhanden sind, empfehlen wir die Verwendung möglichst wenig-informativer Verteilungen. Für die Varianz wären dies zum Beispiel die Verteilungen $U(0, \infty)$ oder $HN(0, 10^6)$. Sollten nur wenige historische Studien verfügbar sein, empfiehlt es sich schwach-informative Verteilungen zu benutzen, deren jeweilige Wahrscheinlichkeitsmasse auf einem möglichst realistischen Wertebereich der Zielgröße liegt. In den hier analysierten Studiengabeheiten ist die Half-Normal-Verteilung $HN(0, 1)$ die favorisierte Startverteilung. In den Simulationsszenarien, in denen 5 oder 6 historische Studien pro Kollektiv existieren, nähern sich die Schätzungen mit $HN(0, 1)$ am schnellsten dem wahren Wert an. Darüber hinaus werden die wenigsten Ausreißer generiert. Auch beim Vorliegen von nur 3 oder 4 Studien liefert die Half-Normal-Verteilung bereits robuste Schätzungen. Insgesamt verhält sie sich von den betrachteten Startverteilungen sowohl in den simulierten Szenarien als auch in

den klinischen Beispielen am besten. Als Alternative bietet sich die Half-Cauchy-Verteilung $HC(0,1)$ an. Die beschränkte Gleichverteilung $U(0,100)$ ist für die Zielgrößen wie das Log Odds Ratio oder die standardisierte Mittelwertdifferenz zu konservativ, wenn nur wenige Beobachtungen vorliegen. Anfänglich wurde die $\Gamma^{-1}(0.001,0.001)$ -Verteilung als übliche a-priori Verteilung für den Varianzparameter genutzt [60, 61]. Inzwischen wurde festgestellt, dass sie in speziellen Situationen zu nicht properen a-posteriori Verteilungen führt [21]. Von der Verwendung der Inverse-Gamma-Verteilung $\Gamma(\epsilon, \epsilon)$ mit $\epsilon \rightarrow 0$ raten wir deshalb ab. Obwohl diese Verteilung als nicht-informativ gilt, haben wir gesehen, dass die Varianz oftmals zu niedrig geschätzt wird. Infolgedessen verhält sie sich antikonservativ und in diesem Zusammenhang höchst informativ.

Auf die Frage nach der notwendigen Mindestanzahl historischer Studien können wir nur eine annähernde Antwort geben. Die Anzahl ist immer von der zu untersuchenden Fragestellung abhängig. Selbstverständlich ist es am besten so viele Studien wie möglich zu betrachten und in die Analysen einzuschließen. Wir empfehlen aufgrund unsere Untersuchungen, dass bei einem Kollektiv mindestens 5 historische Studien vorhanden sein sollten. Nur in diesem Fall ist gewährleistet, dass die einzelnen a-posteriori Verteilungen mehr durch die Daten als durch die gewählten a-priori Verteilungen geprägt sind.

Unter Berücksichtigung dieser Erkenntnisse müssen wir die Interpretation der Resultate für das FeNO-Beispiel 4.3 neu überdenken. Da in dem Beispiel nur zwei frühere Erwachsenenstudien respektive drei Kinderstudien vorhanden sind, kann der Behandlungseffekt einer zukünftigen Studie nicht sicher vorhergesagt werden. Die Studien- und Fallzahlen sind im FeNO-Beispiel zu gering, um eine korrekte Prognose für den zukünftigen Behandlungseffekt zuzulassen. Daher sind die Resultate stark durch die gewählte a-priori Verteilung für den Interstudienvarianzparameter geprägt. Somit könnten die Ergebnisse der Kinder zwar homogen, aber eventuell verfälscht sein. Andererseits könnten auch die gegenläufigen Studienergebnisse der Erwachsenen die wahre Therapiewirkung suggerieren.

Die in dieser Arbeit neu vorgestellte α -Trimming-Methode der Markovketten stellt ein geeignetes Hilfsmittel zur Überprüfung der Stabilität der Parameterschätzungen dar. Da die Markovketten jedoch durch die α -Trimming-Methode datengesteuert beschränkt werden, ist die Verwendung der α -getrimmten Parameter als tatsächliche Planungsgrundlage bzw. informative a-priori Parameter für eine neue Studie allerdings nicht zulässig (vgl. Kap. 3.6). Mit

dieser Technik lässt sich aber gut kontrollieren, inwiefern die Schätzungen sensitiv auf die gewählten hyper a-priori Verteilungen des hierarchischen Modells reagieren.¹

Der hier detailliert analysierte meta-analytisch prädiktive Ansatz berücksichtigt bereits vorhandene historische Daten. Dadurch können in neuen Studien Patienten eingespart werden. Weil deshalb die Rekrutierung schneller abgeschlossen werden kann, führt dies zu einer Beschleunigung der klinischen Studie und somit zu früher vorliegenden klinischen Erkenntnissen. Dadurch wird die neue Studie machbarer und unter anderem ethisch leichter zu rechtfertigen. Wie viele Patienten tatsächlich in einer neuen Studie eingespart werden können, hängt aber nicht nur von der Qualität der historischen Ergebnisse und der mathematisch ermittelten virtuellen Fallzahleinsparung N_{prior} ab. Eine Vielzahl unterschiedlicher Faktoren wie etwa das Allokationsschema, das nicht unbedingt auf gleich große Gruppengrößen ausgerichtet sein muss, oder sekundäre Zielgrößen und explorative Fragestellungen beeinflussen die benötigte Menge an neuen Patienten. Darüber hinaus müssen die Studienplaner diesem Bayes-Verfahren genügend Vertrauen bezüglich der Beurteilungsfähigkeit der historischen Studienresultate entgegenbringen, um eine Anwendung überhaupt realisierbar zu machen. Die Planung und die Antragsstellung einer neuen Studie werden mit diesem Bayes-Verfahren komplizierter. Die Verantwortlichen müssen die verwendeten Methoden detailliert beschreiben und gegenüber den Behörden rechtfertigen, auf welcher Grundlage sie genau diese eine informative a-priori Verteilung, die das Vorwissen repräsentiert, in die Auswertung der neuen Studie einbringen. Dieser zunächst nachteilig wirkende Aspekt ist aber nicht nur auf den meta-analytisch prädiktiven Ansatz beschränkt, sondern ist bei allen komplexeren Studiendesigns gegeben. Trotzdem überwiegen die Vorteile der Methodik. Der Informationsgehalt der historischen Studien wird objektiv ermittelt und beruht nicht auf Willkür oder Expertenwissen alleine. Die neue Studie wird durch die Einsparung an neu zu rekrutierenden Patienten umsetzbarer, sodass dies den Mehraufwand in der Planung kompensiert. Die Flexibilität der hierarchischen Modelle ermöglicht es die historischen Ergebnisse der Kontroll- und der experimentellen Behandlungsgruppe verschiedener Kollektive gemeinsam zu extrapolieren. Andererseits ermöglichen die Modellierungstechniken des Bayes-Verfahrens zudem auch nur Verwendung der historischen Kontrollen. In diesem Fall bietet sich die Anwendung des meta-analytisch prädiktiven Ansatzes insbesondere für Phase-IIa-Studien an, wo häufig keine direkte Kontrollgruppe verfügbar ist.

¹ Einzelne Ergebnisse dieser Dissertation wurden bereits in *Koch et al.* [36–39] präsentiert.

Die Rückschlüsse, die wir aus dieser Arbeit ziehen können, weisen aber auch ein paar Limitationen auf. Wir haben insgesamt lediglich vier ausgewählte a-priori Verteilungen mit jeweils einer Parameterkombination in allgemeineren Studiensituationen untersucht. Bei einer anderen Auswahl der Parameter ändern sich das Verhalten der Verteilungen und die Auswirkungen auf die Parameterschätzungen. Was für eine Zielgröße als schwach-informativ gilt, kann in einem anderen Kontext bereits hoch-informativ sein. Wenn wir beispielsweise als Zielgröße ein Log Odds Ratio betrachten, ist die $HN(0,1)$ -Verteilung als a-priori Verteilung für die Standardabweichung angemessen. Für eine Zielgröße wie zum Beispiel ein Mittelwertunterschied im systolischen Blutdruck in mmHG, der realistisch zwischen 100-225 mmHG schwanken kann, ist die $HN(0,1)$ -Verteilung viel zu informativ und in dieser Parameterkombination vollkommen unpassend. Die Varianz der Half-Normal-Verteilung muss hierfür größer gewählt werden. Demzufolge ist die unabdingbare Forderung, die a-priori Verteilungen immer geeignet an die jeweilige Fragestellung und Zielgröße anzupassen. Die Bestimmung der a-posteriori prädiktiven Verteilung erfolgte in dieser Arbeit mittels MCMC-Verfahren. Alternative Berechnungsmethoden der hierarchischen Modelle, beispielsweise Integrated Nested Laplace Approximations (INLA)[54], könnten eventuell robustere Parameterschätzungen ergeben. Diese Untersuchung stellt einen zukünftigen Forschungsaspekt dar.

Eine weitere Einschränkung ist, dass wir den meta-analytisch prädiktiven Ansatz explizit nur für normalverteilte Zielgrößen untersucht haben. Für den Nachweis der Alternativhypothese in der neuen Studie haben wir ein Normal-Normal-Modell zugrunde gelegt. Dieses Modell besitzt eine sehr einfache Form. Ein Forschungsziel ist es darum das meta-analytisch prädiktive Verfahren auf komplexere Modelle wie etwa nicht-parametrische Situationen zu übertragen. Dafür muss das hierarchische Modell angepasst werden. Wie im nicht-parametrischen Fall aus den prädiktiven Parametern eine virtuelle Fallzahl, egal ob N_{eff} oder N_{prior} explizit berechnet werden kann, ist offen. Weiter ist es erstrebenswert, eine ähnliche Methode für Überlebenszeiten zu entwickeln. Eine bayesianische Metaanalyse für kumulative Überlebenskurven ist nicht trivial. Selbst in der klassischen Statistik ist eine Random-Effects-Metaanalyse für zensierte Zielgrößen kompliziert [5]. Jede historische Kaplan-Meier-Kurve muss entsprechend modelliert werden. Der Umgang mit zensierten Daten und die Voraussetzung, dass die historischen Ereignis- und Zensierungszahlen im Zeitverlauf vorliegen müssen, erschweren die Modellanpassung. Eine Übertragung der Methodik auf Hazardraten bzw. Hazard Ratios scheint jedoch möglich und vergleichsweise einfach umsetzbar zu sein. In einer aktuellen Forschungsarbeit ist die prädiktive Methode von *Gsteiger et al.* [30] bereits für die Anwendung auf Zähldaten mit Overdispersion erweitert worden. Für diese Daten wird ein negativ binomiales hierarchisches

Modell angepasst. Allerdings beschränken sich *Gsteiger et al.* [30] auf die Verwendung der historischen Informationen der Kontrollgruppe und auf die Betrachtung von nur einem Patientenkollektiv. Ebenfalls wird eine Fallzahl für den Wert der historischen Daten berechnet, die jedoch genauso wie N_{eff} in *Neuenschwander et al.* [50] definiert wird. Die Fallzahl der Kontrollgruppe einer neuen Studie wird um diese Anzahl virtueller Patienten reduziert. Wie wir gezeigt haben, entspricht N_{eff} aber weder der a-priori Information noch der wirklichen Anzahl einsparbarer Patienten N_{prior} .

Letztendlich stellt der in dieser Arbeit vorgestellte meta-analytisch prädiktive Ansatz eine geeignete Option dar, um Patienten in einer zukünftigen Studie einzusparen. Die Weiterentwicklung von *Gerß et al.* [24] erlaubt es, Daten aus verschiedenen Kollektiven miteinander zu kombinieren. In der klinischen Praxis werden Behandlungsstrategien häufig zur Genüge an Erwachsenen untersucht. Diese Erfahrungen können dazu genutzt werden, den Behandlungserfolg bei Kindern zu schätzen. Die Voraussetzungen sind, dass die Patientenkollektive austauschbar und möglichst gleich aufgebaut sind. Um einen systematischen Therapieunterschied zwischen erwachsenen und pädiatrischen Patienten aufzudecken, muss in beiden Kollektiven eine identische Fragestellung betrachtet werden. Der Umfang der Extrapolation wird maßgeblich durch die Heterogenität der historischen Ergebnisse bestimmt. Wie wir im klinischen Beispiel 4.1 gesehen haben, kann dort die Fallzahl einer zukünftigen Kinderstudie um ca. 10% reduziert werden. In den Simulationsszenarien bei adjustiertem Erwartungswert beträgt die Einsparung sogar bis zu 100%, sodass die Durchführung einer neuen Studie zu dieser Fragestellung nicht mehr notwendig ist. Allerdings sollte von Fall zu Fall neu entschieden werden, wie viele Patienten tatsächlich in der zukünftigen Studie neu rekrutiert werden.

Alles in allem ist der meta-analytisch prädiktive Ansatz eine geeignete und flexible Methode, um den Informationsgehalt der historischen Studien objektiv zu beurteilen und die informativen a-priori Parameter einer neuen Studie empirisch zu bestimmen. Das Verfahren ermöglicht daher eine rationale Auswahl der informativen Startparameter und erfüllt somit die essentielle Forderung von *J.M. Bernardo* [34, S. 12]:

In the complex multiparameter models [...] there is little hope for a detailed assessment of a huge personal multivariate prior; the naive use of some “tractable” prior may then hide important unwarranted assumptions which may easily dominate the analysis [...]. Careful, responsible choice of a non-subjective prior is possibly the best available alternative.

Appendix

A Gegenbeispiel Proportionalität von Fallzahl und Präzision

Beispiel A.1. An dieser Stelle wollen wir ein zusätzliches Gegenbeispiel angeben, dass die Annahme der Proportionalität von Fallzahl und Präzision in *Neuenschwander et al.* [50] widerlegt. Selbst im einfachen hierarchischen Normal-Normal-Modell gilt diese Beziehung nicht. In Beispiel 3.1 haben wir schon eine Situation betrachtet, in der wir zeigen konnten, dass die Annahme in Gleichung (3.9) im Allgemeinen nicht gilt. Dafür haben wir angenommen, dass nur eine historische Studie existiert. Allerdings könnte mancher Leser denken, dass in dieser Situation dann auch die Interstudienvarianz $\tau = 0$ ist. Um diese Argumentation zusätzlich zu entkräften, betrachten wir das folgende Beispiel. Dafür benötigen wir ein paar ergänzende Voraussetzungen. Wie in Beispiel 3.1 setzen wir voraus, dass nur ein historisches Kollektiv existiert. Die Zielgröße sei das Log Odds Ratio. Wir nehmen an, dass 2 historische Studien $h = 1, 2$ existieren, bei denen die Gruppengrößen jeweils gleich sind

$$n_{T,1} = n_{T,2} = n_{C,1} = n_{C,2} = N_{hist}/4.$$

Die quadrierten Intrastudienstandardfehler σ_1^2, σ_2^2 seien ebenfalls identisch

$$\begin{aligned} \sigma_1^2 &= \frac{1}{N_{hist}} \left(\frac{2}{p_{T,1}(1-p_{T,1})} + \frac{2}{p_{C,1}(1-p_{C,1})} \right) \\ &= \frac{1}{N_{hist}} \cdot s_1^2 = \frac{1}{N_{hist}} \cdot s_2^2 \\ &= \frac{1}{N_{hist}} \left(\frac{2}{p_{T,2}(1-p_{T,2})} + \frac{2}{p_{C,2}(1-p_{C,2})} \right) = \sigma_2^2, \end{aligned}$$

wobei s_1, s_2 die empirischen Interstudienstandardabweichungen seien. Das zugrundeliegende Modell sei das hierarchische Normal-Normal-Modell aus Kapitel 2.4.2 mit bekannter Interstudienvarianz $\tau^2 \neq 0$. Die a-posteriori prädiktive Varianz V_τ hat damit die Gestalt in Gleichung (2.40). Somit folgt

$$\begin{aligned} V_\tau &= \frac{1}{\sum_{h=1}^2 w_h} + \tau^2 = \frac{1}{\frac{1}{\sigma_1^2 + \tau^2} + \frac{1}{\sigma_2^2 + \tau^2}} + \tau^2 \\ &= \frac{1}{\frac{2}{\sigma_1^2 + \tau^2}} + \tau^2 = \frac{1}{2}(\sigma_1^2 + \tau^2) + \tau^2 \end{aligned} \tag{A.1}$$

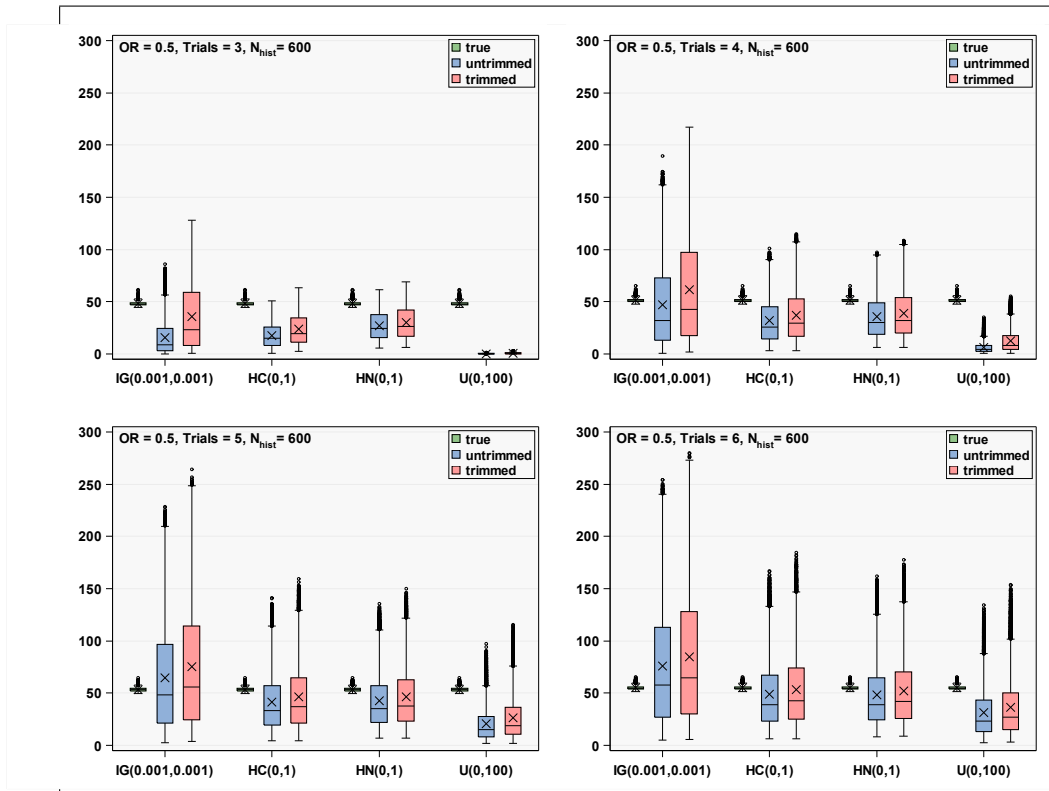
$$\begin{aligned} &= \frac{1}{2}\sigma_1^2 + \frac{3}{2}\tau^2 = \frac{1}{N_{hist}} \cdot \frac{1}{2}s_1^2 + \frac{3}{2}\underbrace{\tau^2}_{\neq 0} \\ &\neq c \cdot \frac{1}{N_{hist}} \end{aligned}$$

mit einer Konstanten $c \in \mathbb{R}$. Somit steht die Fallzahl im Normal-Normal-Modell nicht in einer proportionalen Beziehung zur a-posteriori prädiktiven Präzision. Die Herleitung der Precision Weight Effective Sample Size N_{eff} ist daher nur intuitiv.

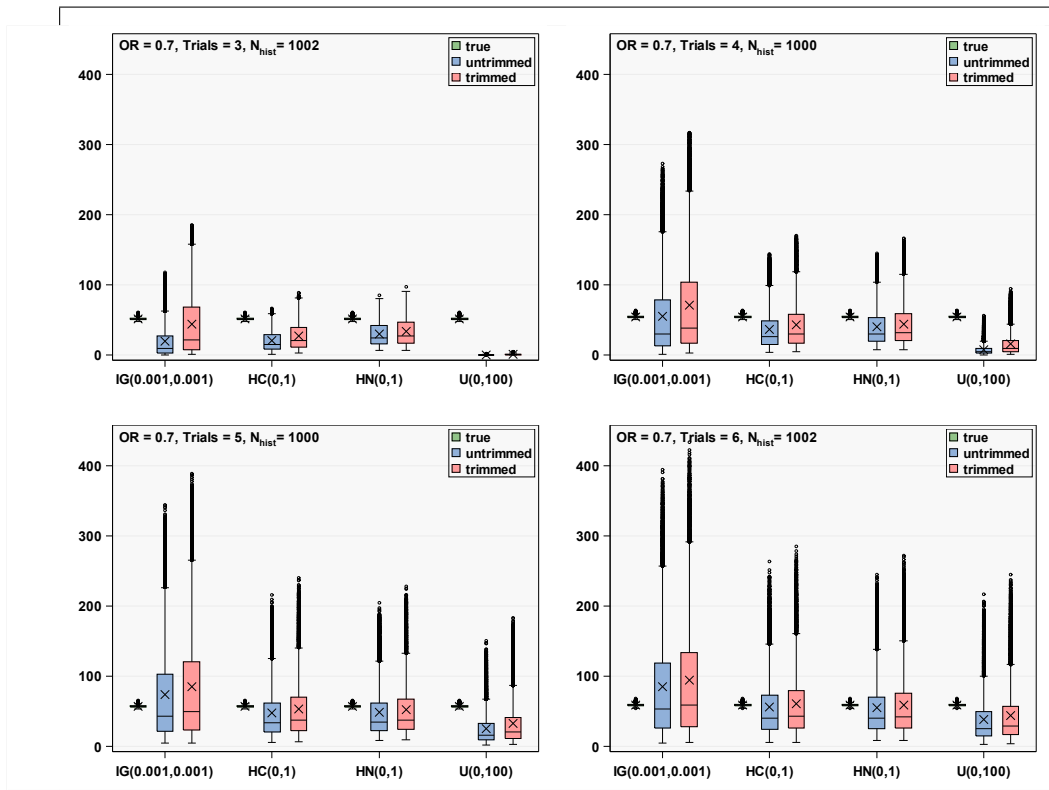
B Anhänge zur Simulationsstudie

B.1 Deskriptive Tabellen und Grafiken der Ergebnisse

B.1.1. Ergebnisse der nicht adjustierten Simulationen



(a) Szenario A: $OR = 0.5$.



(b) Szenario B: $OR = 0.7$.

Abb. B.1: Boxplots der Precision Weight Effective Sample Size N_{eff} .

Number of historical trials	Prior distribution for τ		Standard	10%	25%	75%	90%	Maximum	Skewness			
			Mean	Deviation	Quantil	Quantil	Median			Quantil	Quantil	
3	$\Gamma^{-1}(0.001, 0.001)$	N_{true}	48.3	1.8	44.8	46.3	47.0	47.9	49.2	50.7	61.4	1.1
		N_{eff}	16.0	16.5	0.0	1.6	3.3	9.0	24.7	42.1	86.0	1.3
		N_{eff_trim}	36.1	33.0	0.9	4.1	8.1	23.3	58.8	90.5	127.9	0.9
	HC(0,1)	N_{true}	48.3	1.8	44.8	46.3	47.0	47.9	49.2	50.7	61.4	1.1
		N_{eff}	17.7	10.9	0.5	5.6	8.5	14.9	25.6	34.5	50.6	0.7
		N_{eff_trim}	23.7	14.9	2.5	7.2	11.1	19.7	34.7	47.4	63.5	0.6
	HN(0,1)	N_{true}	48.3	1.8	44.8	46.3	47.0	47.9	49.2	50.7	61.4	1.1
		N_{eff}	27.2	13.5	6.0	11.6	15.8	24.3	37.7	48.1	61.8	0.5
		N_{eff_trim}	29.9	15.4	6.3	12.3	16.8	26.3	41.8	53.9	69.3	0.6
	U(0,100)	N_{true}	48.3	1.8	44.8	46.3	47.0	47.9	49.2	50.7	61.4	1.1
		N_{eff}	0.3	0.2	0.1	0.1	0.2	0.3	0.4	0.6	1.1	0.9
		N_{eff_trim}	0.7	0.5	0.1	0.2	0.3	0.5	1.0	1.6	2.8	1.1
4	$\Gamma^{-1}(0.001, 0.001)$	N_{true}	51.4	1.8	47.6	49.4	50.1	51.1	52.4	53.7	65.6	1.0
		N_{eff}	47.3	41.2	0.6	7.5	13.4	32.2	72.8	113.2	189.8	1.0
		N_{eff_trim}	61.8	52.8	2.2	9.6	17.3	43.0	97.3	146.7	217.3	0.9
	HC(0,1)	N_{true}	51.4	1.8	47.6	49.4	50.1	51.1	52.4	53.7	65.6	1.0
		N_{eff}	31.8	21.0	2.9	10.0	14.6	25.6	45.0	64.5	101.0	0.9
		N_{eff_trim}	37.2	25.2	3.3	11.3	16.7	29.6	53.1	76.5	114.6	0.9
	HN(0,1)	N_{true}	51.4	1.8	47.6	49.4	50.1	51.1	52.4	53.7	65.6	1.0
		N_{eff}	35.8	20.7	6.1	13.9	18.9	30.0	49.2	67.7	97.3	0.8
		N_{eff_trim}	39.1	23.5	6.3	14.6	20.0	32.3	54.0	75.7	108.3	0.9
	U(0,100)	N_{true}	51.4	1.8	47.6	49.4	50.1	51.1	52.4	53.7	65.6	1.0
		N_{eff}	6.1	5.4	0.4	1.5	2.3	4.2	8.2	13.9	35.4	1.7
		N_{eff_trim}	12.5	11.0	0.7	2.7	4.3	8.4	17.8	29.7	55.4	1.3
5	$\Gamma^{-1}(0.001, 0.001)$	N_{true}	53.5	1.7	49.6	51.5	52.2	53.2	54.5	55.8	64.4	0.9
		N_{eff}	64.8	52.1	2.5	12.1	21.5	48.1	96.9	146.4	228.6	1.0
		N_{eff_trim}	75.5	60.6	3.7	13.6	24.6	56.0	114.2	170.6	264.3	0.9
	HC(0,1)	N_{true}	53.5	1.7	49.6	51.5	52.2	53.2	54.5	55.8	64.4	0.9
		N_{eff}	41.6	27.8	4.2	13.0	19.5	33.6	57.4	83.4	141.2	1.0
		N_{eff_trim}	46.5	31.6	4.6	14.1	21.4	37.1	64.5	94.4	159.1	1.0
	HN(0,1)	N_{true}	53.5	1.7	49.6	51.5	52.2	53.2	54.5	55.8	64.4	0.9
		N_{eff}	42.6	26.1	6.9	15.6	21.9	35.0	57.4	82.1	135.8	1.0
		N_{eff_trim}	46.2	29.2	7.1	16.3	23.2	37.6	62.8	90.6	149.8	1.0
	U(0,100)	N_{true}	53.5	1.7	49.6	51.5	52.2	53.2	54.5	55.8	64.4	0.9
		N_{eff}	20.5	16.7	1.6	5.3	8.2	14.9	27.9	45.2	97.6	1.5
		N_{eff_trim}	26.7	21.7	2.0	6.7	10.5	19.2	36.7	59.0	115.2	1.4
6	$\Gamma^{-1}(0.001, 0.001)$	N_{true}	55.1	1.8	51.3	53.0	53.8	54.9	56.1	57.5	65.6	0.8
		N_{eff}	75.7	58.0	5.2	15.6	27.3	57.9	112.8	167.0	254.2	0.9
		N_{eff_trim}	84.9	65.1	5.6	17.0	30.1	64.9	127.9	187.9	279.7	0.8
	HC(0,1)	N_{true}	55.1	1.8	51.3	53.0	53.8	54.9	56.1	57.5	65.6	0.8
		N_{eff}	48.7	32.3	6.2	15.6	23.2	39.3	67.2	98.1	166.8	1.0
		N_{eff_trim}	53.4	35.9	6.6	16.7	25.0	42.7	73.8	108.5	184.4	1.0
	HN(0,1)	N_{true}	55.1	1.8	51.3	53.0	53.8	54.9	56.1	57.5	65.6	0.8
		N_{eff}	48.1	30.2	8.5	17.3	24.5	39.1	64.9	94.3	162.2	1.1
		N_{eff_trim}	51.9	33.3	8.8	18.1	25.8	41.8	70.4	102.9	177.6	1.1
	U(0,100)	N_{true}	55.1	1.8	51.3	53.0	53.8	54.9	56.1	57.5	65.6	0.8
		N_{eff}	31.7	24.2	2.6	8.9	13.5	23.5	43.2	68.2	134.5	1.3
		N_{eff_trim}	36.6	28.2	3.1	9.9	15.3	27.0	50.0	79.5	153.5	1.3

Tab. B.1: Deskriptive Statistiken für die Precision Weight Effective Sample Size N_{eff} : Szenario A, $OR = 0.5$.

Number of historical trials	Prior distribution for τ		Mean	Standard Deviation	Minimum	10% Quantil	25% Quantil	Median	75% Quantil	90% Quantil	Maximum	Skewness
3	$\Gamma^{-1}(0.001, 0.001)$	N_{true}	51.3	1.6	47.4	49.5	50.2	51.1	52.2	53.5	60.5	0.9
		N_{eff}	19.3	22.9	0.0	1.7	3.3	9.0	27.1	55.2	118.0	1.7
		N_{eff_trim}	43.8	46.7	1.1	4.3	7.9	22.0	67.8	124.7	184.8	1.2
	HC(0,1)	N_{true}	51.3	1.6	47.4	49.5	50.2	51.1	52.2	53.5	60.5	0.9
		N_{eff}	20.2	14.2	0.6	6.0	8.9	15.4	28.8	43.3	66.1	1.0
		N_{eff_trim}	27.3	19.8	2.8	7.7	11.4	20.3	39.3	60.2	88.9	1.0
	HN(0,1)	N_{true}	51.3	1.6	47.4	49.5	50.2	51.1	52.2	53.5	60.5	0.9
		N_{eff}	30.4	17.5	6.5	12.2	16.0	24.8	42.0	59.1	85.3	0.8
		N_{eff_trim}	33.6	20.4	6.8	12.9	17.1	26.8	46.5	67.3	97.1	0.9
	U(0,100)	N_{true}	51.3	1.6	47.4	49.5	50.2	51.1	52.2	53.5	60.5	0.9
		N_{eff}	0.4	0.2	0.1	0.1	0.2	0.3	0.5	0.7	1.6	1.2
		N_{eff_trim}	0.9	0.8	0.1	0.2	0.3	0.6	1.2	2.1	4.7	1.5
4	$\Gamma^{-1}(0.001, 0.001)$	N_{true}	54.7	1.6	50.6	52.8	53.5	54.5	55.6	56.8	63.4	0.8
		N_{eff}	54.8	57.2	1.2	8.1	13.2	29.7	78.3	147.3	272.4	1.5
		N_{eff_trim}	71.3	73.6	3.2	10.3	16.9	38.6	103.7	192.8	316.5	1.4
	HC(0,1)	N_{true}	54.7	1.6	50.6	52.8	53.5	54.5	55.6	56.8	63.4	0.8
		N_{eff}	36.4	28.6	3.9	10.7	15.3	25.8	48.9	81.3	143.4	1.4
		N_{eff_trim}	42.8	34.7	4.7	12.0	17.3	29.7	57.8	97.6	170.1	1.4
	HN(0,1)	N_{true}	54.7	1.6	50.6	52.8	53.5	54.5	55.6	56.8	63.4	0.8
		N_{eff}	40.3	28.1	7.5	14.7	19.3	30.0	53.2	84.6	144.7	1.3
		N_{eff_trim}	44.3	32.4	7.8	15.4	20.5	32.3	58.4	95.1	166.5	1.4
	U(0,100)	N_{true}	54.7	1.6	50.6	52.8	53.5	54.5	55.6	56.8	63.4	0.8
		N_{eff}	7.6	8.0	0.5	1.7	2.5	4.5	9.4	18.4	55.8	2.2
		N_{eff_trim}	15.8	16.6	1.0	3.1	4.7	9.1	20.3	40.5	94.0	1.9
5	$\Gamma^{-1}(0.001, 0.001)$	N_{true}	57.0	1.6	53.1	55.1	55.8	56.8	57.9	59.1	65.6	0.7
		N_{eff}	73.6	71.8	4.6	13.1	21.1	43.5	103.2	188.0	344.0	1.5
		N_{eff_trim}	85.3	83.5	5.2	14.7	23.8	49.6	120.3	220.2	388.1	1.4
	HC(0,1)	N_{true}	57.0	1.6	53.1	55.1	55.8	56.8	57.9	59.1	65.6	0.7
		N_{eff}	47.7	38.1	6.0	14.3	20.4	33.6	62.1	105.1	216.0	1.5
		N_{eff_trim}	53.4	43.6	6.5	15.4	22.3	37.0	69.7	118.8	240.2	1.6
	HN(0,1)	N_{true}	57.0	1.6	53.1	55.1	55.8	56.8	57.9	59.1	65.6	0.7
		N_{eff}	48.3	36.0	8.6	16.7	22.7	35.0	62.1	102.9	204.9	1.5
		N_{eff_trim}	52.7	40.6	9.0	17.6	23.9	37.4	67.5	113.8	227.4	1.6
	U(0,100)	N_{true}	57.0	1.6	53.1	55.1	55.8	56.8	57.9	59.1	65.6	0.7
		N_{eff}	25.4	24.3	1.9	6.1	9.1	16.0	32.3	59.6	149.9	1.9
		N_{eff_trim}	32.9	31.7	2.8	7.7	11.6	20.5	41.6	78.4	183.2	1.9
6	$\Gamma^{-1}(0.001, 0.001)$	N_{true}	58.6	1.6	54.2	56.7	57.5	58.5	59.6	60.7	67.8	0.7
		N_{eff}	84.7	79.1	4.8	16.8	26.2	53.0	118.4	210.6	393.5	1.4
		N_{eff_trim}	94.5	88.8	5.3	18.2	28.5	58.6	133.6	237.9	433.3	1.4
	HC(0,1)	N_{true}	58.6	1.6	54.2	56.7	57.5	58.5	59.6	60.7	67.8	0.7
		N_{eff}	55.8	44.3	5.8	17.1	24.2	39.9	72.9	122.9	263.4	1.6
		N_{eff_trim}	61.0	49.3	6.1	18.2	25.9	43.1	79.7	136.2	284.4	1.6
	HN(0,1)	N_{true}	58.6	1.6	54.2	56.7	57.5	58.5	59.6	60.7	67.8	0.7
		N_{eff}	54.7	41.5	8.1	18.6	25.2	39.9	70.5	117.0	244.3	1.6
		N_{eff_trim}	59.1	46.1	8.3	19.5	26.5	42.4	76.1	128.3	271.7	1.6
	U(0,100)	N_{true}	58.6	1.6	54.2	56.7	57.5	58.5	59.6	60.7	67.8	0.7
		N_{eff}	38.5	34.6	3.2	10.2	14.9	25.6	49.1	88.9	216.3	1.8
		N_{eff_trim}	44.4	40.4	3.5	11.5	16.9	29.1	56.8	103.3	244.4	1.8

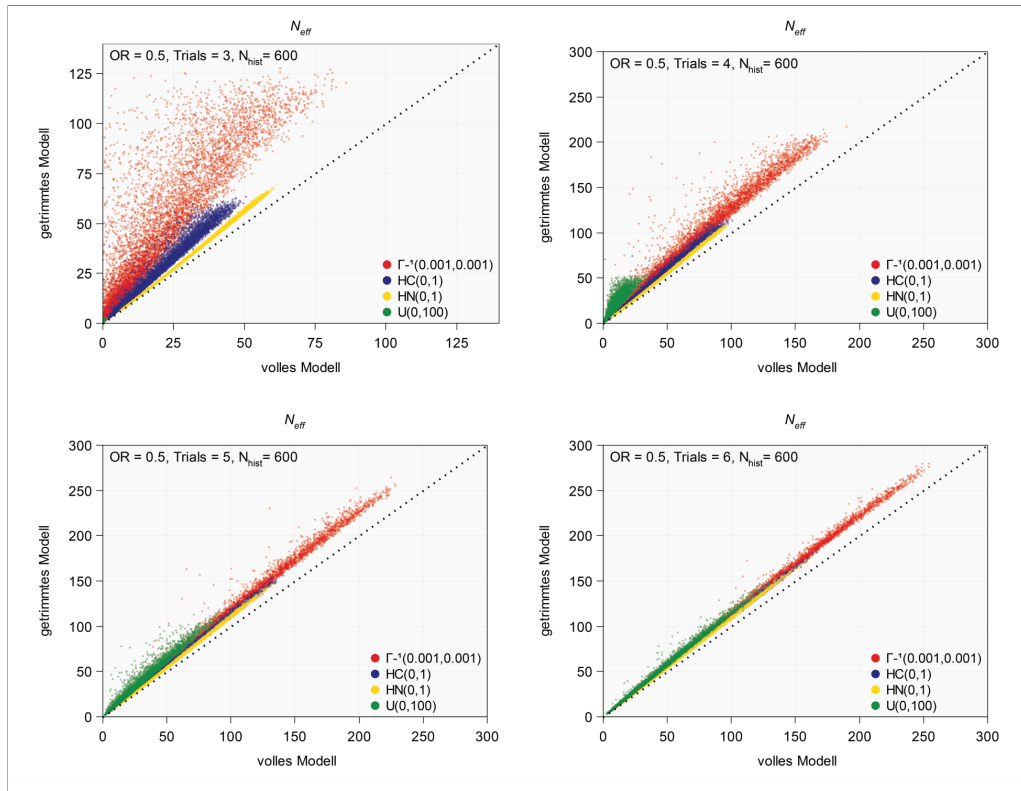
Tab. B.2: Deskriptive Statistiken für die Precision Weight Effective Sample Size N_{eff} : Szenario B, $OR = 0.7$.

Number of historical trials	Prior distribution for τ		Standard		10%	25%	75%	90%	Maximum	Skewness		
			Mean	Deviation	Minimum	Quantil	Quantil	Median			Quantil	Quantil
3	$\Gamma^{-1}(0.001,0.001)$	$N_{\text{eff_trinf}}/N_{\text{eff}}$	3.39	15.71	1.28	1.66	1.83	2.14	2.77	4.19	1352.96	66.36
	$HC(0,1)$	$N_{\text{eff_trinf}}/N_{\text{eff}}$	1.34	0.47	1.13	1.21	1.24	1.29	1.36	1.47	40.82	62.34
	$HN(0,1)$	$N_{\text{eff_trinf}}/N_{\text{eff}}$	1.09	0.03	1.04	1.05	1.06	1.08	1.11	1.12	1.17	0.35
	$U(0,100)$	$N_{\text{eff_trinf}}/N_{\text{eff}}$	2.04	0.47	1.28	1.53	1.67	1.93	2.33	2.73	4.76	0.96
4	$\Gamma^{-1}(0.001,0.001)$	$N_{\text{eff_trinf}}/N_{\text{eff}}$	1.32	0.52	1.13	1.20	1.23	1.27	1.34	1.43	48.15	73.37
	$HC(0,1)$	$N_{\text{eff_trinf}}/N_{\text{eff}}$	1.16	0.05	1.08	1.12	1.13	1.15	1.18	1.21	3.25	9.70
	$HN(0,1)$	$N_{\text{eff_trinf}}/N_{\text{eff}}$	1.08	0.02	1.03	1.05	1.06	1.08	1.10	1.12	1.18	0.50
	$U(0,100)$	$N_{\text{eff_trinf}}/N_{\text{eff}}$	2.03	0.55	1.32	1.57	1.69	1.89	2.19	2.63	9.34	3.35
5	$\Gamma^{-1}(0.001,0.001)$	$N_{\text{eff_trinf}}/N_{\text{eff}}$	1.16	0.05	1.08	1.12	1.13	1.15	1.18	1.20	3.41	12.38
	$HC(0,1)$	$N_{\text{eff_trinf}}/N_{\text{eff}}$	1.11	0.02	1.06	1.08	1.09	1.11	1.12	1.14	1.40	1.51
	$HN(0,1)$	$N_{\text{eff_trinf}}/N_{\text{eff}}$	1.07	0.02	1.03	1.05	1.06	1.07	1.09	1.10	1.15	0.43
	$U(0,100)$	$N_{\text{eff_trinf}}/N_{\text{eff}}$	1.30	0.18	1.12	1.20	1.22	1.26	1.33	1.42	6.79	11.05
6	$\Gamma^{-1}(0.001,0.001)$	$N_{\text{eff_trinf}}/N_{\text{eff}}$	1.11	0.03	1.06	1.08	1.10	1.11	1.13	1.14	1.53	1.96
	$HC(0,1)$	$N_{\text{eff_trinf}}/N_{\text{eff}}$	1.09	0.02	1.05	1.07	1.07	1.09	1.10	1.11	1.22	0.76
	$HN(0,1)$	$N_{\text{eff_trinf}}/N_{\text{eff}}$	1.07	0.02	1.03	1.05	1.06	1.07	1.08	1.09	1.14	0.33
	$U(0,100)$	$N_{\text{eff_trinf}}/N_{\text{eff}}$	1.15	0.04	1.08	1.11	1.12	1.14	1.16	1.19	2.24	6.86

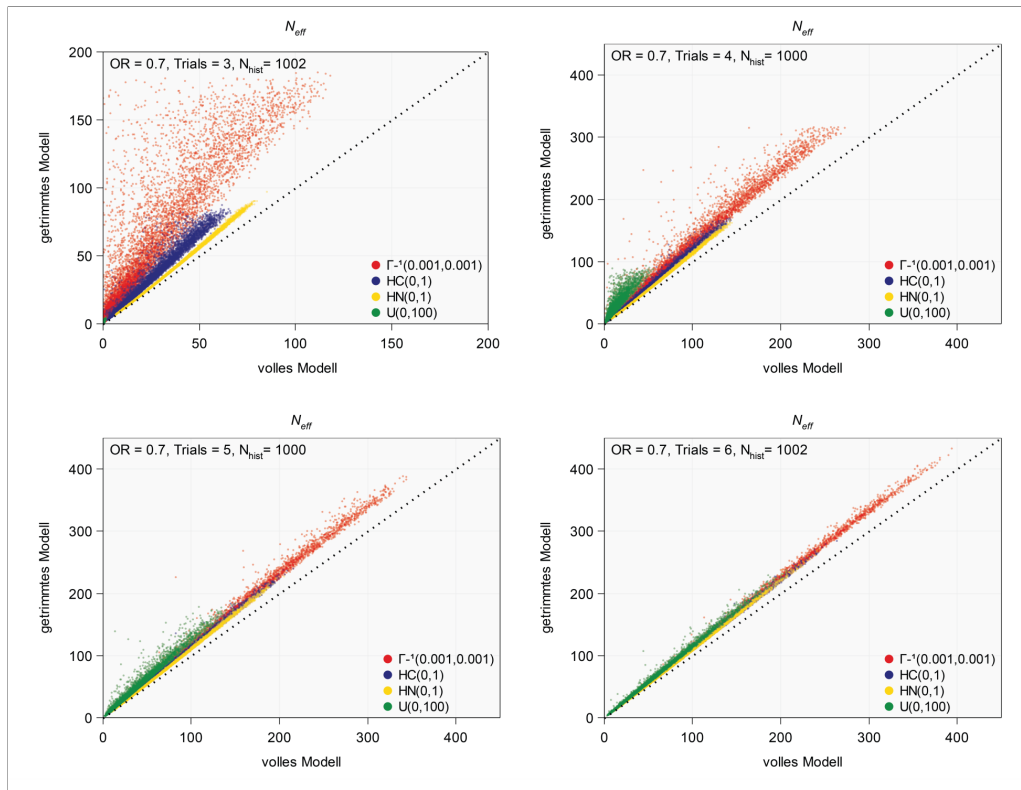
(a) Szenario A: $OR = 0.5$.

Number of historical trials	Prior distribution for τ		Standard		10%	25%	75%	90%	Maximum	Skewness		
			Mean	Deviation	Minimum	Quantil	Quantil	Median			Quantil	Quantil
3	$\Gamma^{-1}(0.001,0.001)$	$N_{\text{eff_trinf}}/N_{\text{eff}}$	3.29	12.43	1.35	1.66	1.83	2.13	2.76	4.20	783.60	42.74
	$HC(0,1)$	$N_{\text{eff_trinf}}/N_{\text{eff}}$	1.35	0.54	1.14	1.21	1.24	1.29	1.37	1.47	42.88	54.18
	$HN(0,1)$	$N_{\text{eff_trinf}}/N_{\text{eff}}$	1.09	0.03	1.04	1.05	1.06	1.08	1.11	1.14	1.22	0.66
	$U(0,100)$	$N_{\text{eff_trinf}}/N_{\text{eff}}$	2.13	0.59	1.28	1.54	1.68	1.95	2.44	2.99	5.46	1.23
4	$\Gamma^{-1}(0.001,0.001)$	$N_{\text{eff_trinf}}/N_{\text{eff}}$	1.31	0.42	1.13	1.19	1.22	1.26	1.33	1.42	25.54	35.92
	$HC(0,1)$	$N_{\text{eff_trinf}}/N_{\text{eff}}$	1.16	0.05	1.08	1.11	1.13	1.15	1.18	1.21	2.26	4.92
	$HN(0,1)$	$N_{\text{eff_trinf}}/N_{\text{eff}}$	1.08	0.03	1.04	1.05	1.06	1.07	1.10	1.12	1.21	0.89
	$U(0,100)$	$N_{\text{eff_trinf}}/N_{\text{eff}}$	2.05	0.65	1.33	1.58	1.69	1.89	2.19	2.63	13.09	5.20
5	$\Gamma^{-1}(0.001,0.001)$	$N_{\text{eff_trinf}}/N_{\text{eff}}$	1.15	0.04	1.07	1.11	1.12	1.14	1.17	1.19	2.75	7.61
	$HC(0,1)$	$N_{\text{eff_trinf}}/N_{\text{eff}}$	1.11	0.03	1.06	1.08	1.09	1.10	1.12	1.14	1.65	2.37
	$HN(0,1)$	$N_{\text{eff_trinf}}/N_{\text{eff}}$	1.07	0.02	1.03	1.05	1.06	1.07	1.09	1.11	1.18	0.84
	$U(0,100)$	$N_{\text{eff_trinf}}/N_{\text{eff}}$	1.29	0.17	1.12	1.19	1.22	1.26	1.31	1.40	10.25	17.84
6	$\Gamma^{-1}(0.001,0.001)$	$N_{\text{eff_trinf}}/N_{\text{eff}}$	1.10	0.02	1.06	1.08	1.09	1.10	1.12	1.13	1.43	1.42
	$HC(0,1)$	$N_{\text{eff_trinf}}/N_{\text{eff}}$	1.08	0.02	1.05	1.06	1.07	1.08	1.09	1.11	1.24	1.10
	$HN(0,1)$	$N_{\text{eff_trinf}}/N_{\text{eff}}$	1.07	0.02	1.03	1.05	1.05	1.06	1.08	1.10	1.18	0.77
	$U(0,100)$	$N_{\text{eff_trinf}}/N_{\text{eff}}$	1.14	0.05	1.08	1.11	1.12	1.13	1.16	1.18	3.90	21.04

(b) Szenario B: $OR = 0.7$.**Tab. B.3:** Deskriptive Statistiken für das Verhältnis von N_{eff} im getrimmten Modell zu N_{eff} im vollen Modell.

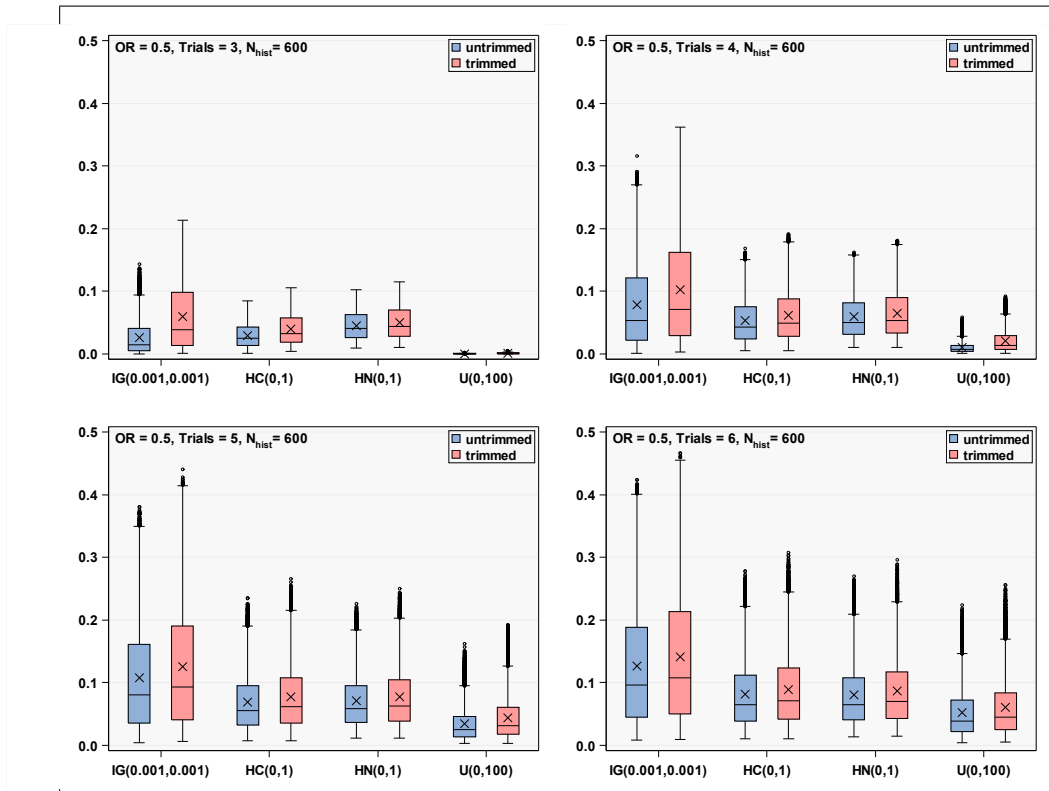


(a) Szenario A: $OR = 0.5$.

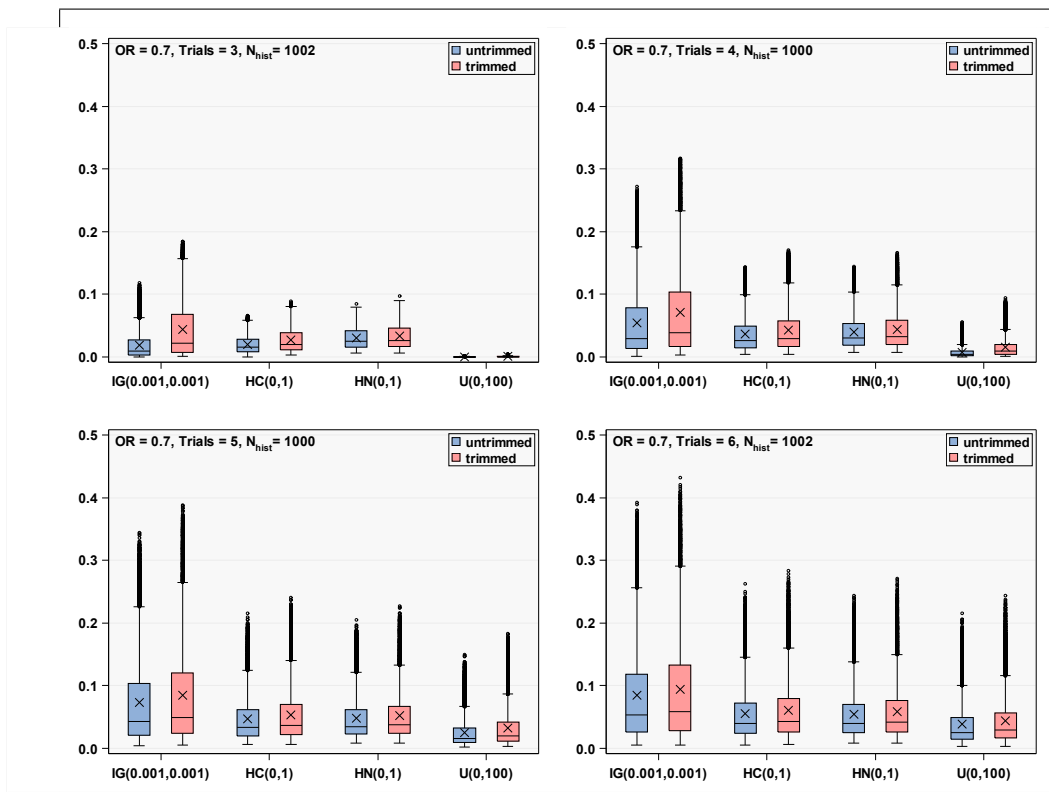


(b) Szenario B: $OR = 0.7$.

Abb. B.2: Scatterplots: Vergleich zwischen N_{eff} im vollen und im getrimmten Modell.



(a) Szenario A: $OR = 0.5$.



(b) Szenario B: $OR = 0.7$.

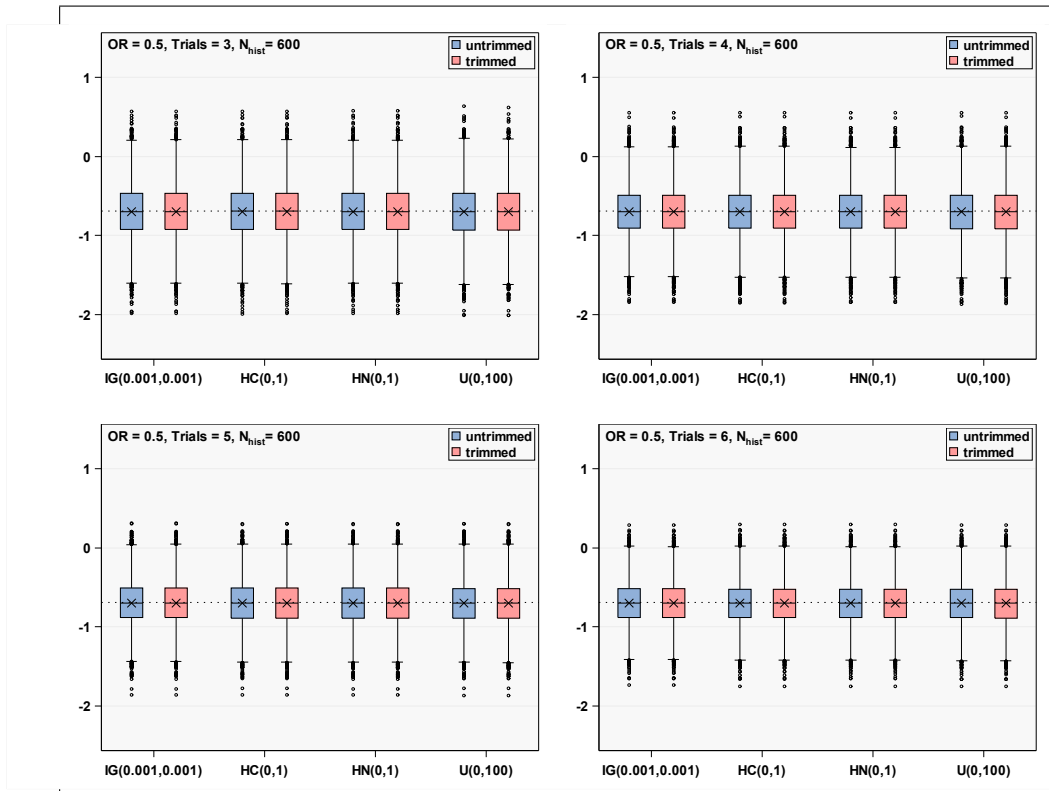
Abb. B.3: Boxplots der Gewichtungsfaktoren ω .

Number of historical trials	Prior distribution for τ		Standard	10%	25%	75%	90%	Maximum	Skewness				
			Mean	Deviation	Quantil	Quantil	Quantil			Quantil			
3	$\Gamma^{-1}(0.001, 0.001)$	ω	0.027	0.027	0.000	0.003	0.006	0.015	0.041	0.070	0.143	1.306	
		ω_{trim}	0.060	0.055	0.002	0.007	0.014	0.039	0.098	0.151	0.213	0.882	
	HC(0,1)	ω	0.030	0.018	0.001	0.009	0.014	0.025	0.043	0.058	0.084	0.671	
		ω_{trim}	0.040	0.025	0.004	0.012	0.018	0.033	0.058	0.079	0.106	0.642	
	HN(0,1)	ω	0.045	0.022	0.010	0.019	0.026	0.040	0.063	0.080	0.103	0.520	
		ω_{trim}	0.050	0.026	0.010	0.020	0.028	0.044	0.070	0.090	0.115	0.555	
	U(0,100)	ω	0.001	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.940	
		ω_{trim}	0.001	0.001	0.000	0.000	0.001	0.001	0.002	0.003	0.005	1.139	
	4	$\Gamma^{-1}(0.001, 0.001)$	ω	0.079	0.069	0.001	0.013	0.022	0.054	0.121	0.189	0.316	0.999
			ω_{trim}	0.103	0.088	0.004	0.016	0.029	0.072	0.162	0.245	0.362	0.899
HC(0,1)		ω	0.053	0.035	0.005	0.017	0.024	0.043	0.075	0.107	0.168	0.891	
		ω_{trim}	0.062	0.042	0.005	0.019	0.028	0.049	0.088	0.128	0.191	0.901	
HN(0,1)		ω	0.060	0.034	0.010	0.023	0.031	0.050	0.082	0.113	0.162	0.838	
		ω_{trim}	0.065	0.039	0.011	0.024	0.033	0.054	0.090	0.126	0.181	0.874	
U(0,100)		ω	0.010	0.009	0.001	0.003	0.004	0.007	0.014	0.023	0.059	1.680	
		ω_{trim}	0.021	0.018	0.001	0.005	0.007	0.014	0.030	0.049	0.092	1.345	
5		$\Gamma^{-1}(0.001, 0.001)$	ω	0.108	0.087	0.004	0.020	0.036	0.080	0.161	0.244	0.381	0.953
			ω_{trim}	0.126	0.101	0.006	0.023	0.041	0.093	0.190	0.284	0.440	0.904
	HC(0,1)	ω	0.069	0.046	0.007	0.022	0.032	0.056	0.096	0.139	0.235	1.023	
		ω_{trim}	0.077	0.053	0.008	0.023	0.036	0.062	0.108	0.157	0.265	1.027	
	HN(0,1)	ω	0.071	0.044	0.011	0.026	0.036	0.058	0.096	0.137	0.226	1.019	
		ω_{trim}	0.077	0.049	0.012	0.027	0.039	0.063	0.105	0.151	0.250	1.043	
	U(0,100)	ω	0.034	0.028	0.003	0.009	0.014	0.025	0.046	0.075	0.163	1.450	
		ω_{trim}	0.045	0.036	0.003	0.011	0.017	0.032	0.061	0.098	0.192	1.381	
	6	$\Gamma^{-1}(0.001, 0.001)$	ω	0.126	0.097	0.009	0.026	0.046	0.097	0.188	0.278	0.424	0.886
			ω_{trim}	0.142	0.108	0.009	0.028	0.050	0.108	0.213	0.313	0.466	0.849
HC(0,1)		ω	0.081	0.054	0.010	0.026	0.039	0.065	0.112	0.163	0.278	1.034	
		ω_{trim}	0.089	0.060	0.011	0.028	0.042	0.071	0.123	0.181	0.307	1.035	
HN(0,1)		ω	0.080	0.050	0.014	0.029	0.041	0.065	0.108	0.157	0.270	1.061	
		ω_{trim}	0.087	0.055	0.015	0.030	0.043	0.070	0.117	0.172	0.296	1.076	
U(0,100)		ω	0.053	0.040	0.004	0.015	0.022	0.039	0.072	0.114	0.224	1.328	
		ω_{trim}	0.061	0.047	0.005	0.017	0.025	0.045	0.083	0.132	0.256	1.314	

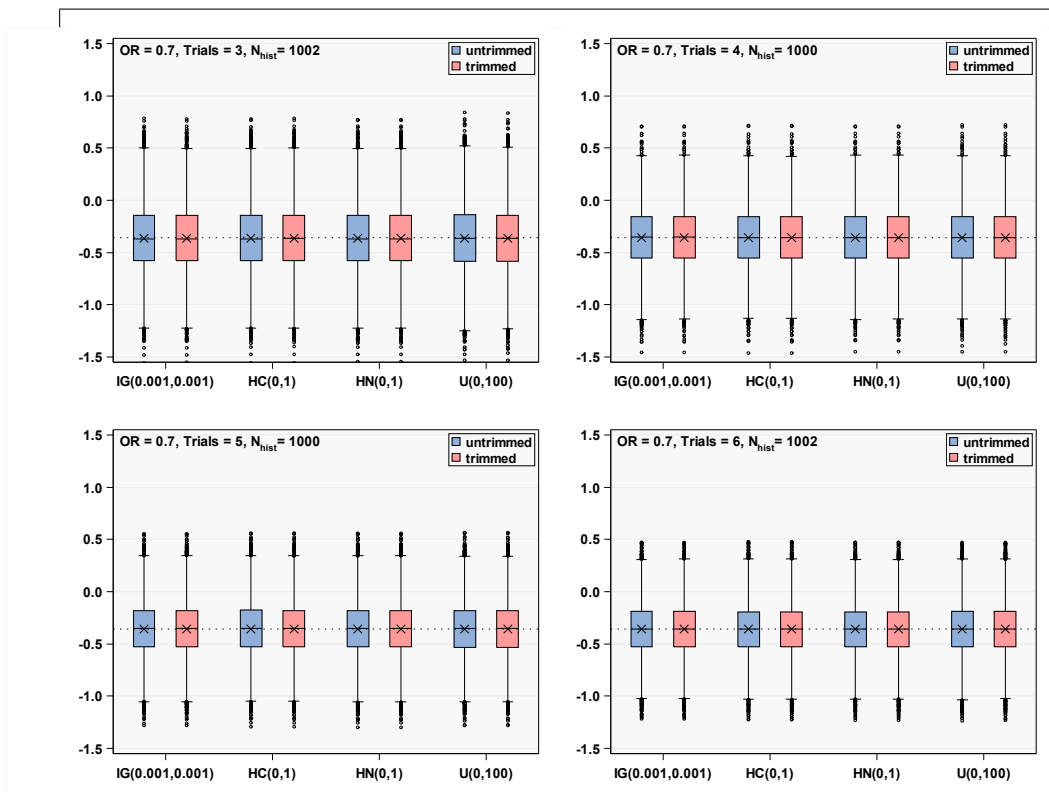
Tab. B.4: Deskriptive Statistiken der Gewichtungsfaktoren ω : Szenario A, $OR = 0.5$.

Number of historical trials	Prior distribution for τ		Mean	Standard Deviation	Minimum	10% Quantil	25% Quantil	Median	75% Quantil	90% Quantil	Maximum	Skewness	
3	$\Gamma^{-1}(0.001, 0.001)$	ω	0.019	0.023	0.000	0.002	0.003	0.009	0.027	0.055	0.118	1.656	
		ω_{trim}	0.044	0.047	0.001	0.004	0.008	0.022	0.068	0.124	0.184	1.208	
	HC(0,1)	ω	0.020	0.014	0.001	0.006	0.009	0.015	0.029	0.043	0.066	0.963	
		ω_{trim}	0.027	0.020	0.003	0.008	0.011	0.020	0.039	0.060	0.089	0.960	
	HN(0,1)	ω	0.030	0.017	0.007	0.012	0.016	0.025	0.042	0.059	0.085	0.830	
		ω_{trim}	0.034	0.020	0.007	0.013	0.017	0.027	0.046	0.067	0.097	0.881	
	U(0,100)	ω	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	1.224	
		ω_{trim}	0.001	0.001	0.000	0.000	0.000	0.001	0.001	0.002	0.005	1.471	
	4	$\Gamma^{-1}(0.001, 0.001)$	ω	0.055	0.057	0.001	0.008	0.013	0.030	0.078	0.147	0.272	1.464
			ω_{trim}	0.071	0.074	0.003	0.010	0.017	0.039	0.104	0.193	0.317	1.374
HC(0,1)		ω	0.036	0.029	0.004	0.011	0.015	0.026	0.049	0.081	0.143	1.370	
		ω_{trim}	0.043	0.035	0.005	0.012	0.017	0.030	0.058	0.098	0.170	1.391	
HN(0,1)		ω	0.040	0.028	0.008	0.015	0.019	0.030	0.053	0.085	0.145	1.327	
		ω_{trim}	0.044	0.032	0.008	0.015	0.020	0.032	0.058	0.095	0.167	1.380	
U(0,100)		ω	0.008	0.008	0.001	0.002	0.003	0.005	0.009	0.018	0.056	2.185	
		ω_{trim}	0.016	0.017	0.001	0.003	0.005	0.009	0.020	0.041	0.094	1.860	
5		$\Gamma^{-1}(0.001, 0.001)$	ω	0.074	0.072	0.005	0.013	0.021	0.043	0.103	0.188	0.344	1.451
			ω_{trim}	0.085	0.083	0.005	0.015	0.024	0.050	0.120	0.220	0.388	1.412
	HC(0,1)	ω	0.048	0.038	0.006	0.014	0.020	0.034	0.062	0.105	0.216	1.532	
		ω_{trim}	0.053	0.044	0.007	0.015	0.022	0.037	0.070	0.119	0.240	1.551	
	HN(0,1)	ω	0.048	0.036	0.009	0.017	0.023	0.035	0.062	0.103	0.205	1.543	
		ω_{trim}	0.053	0.041	0.009	0.018	0.024	0.037	0.067	0.114	0.227	1.582	
	U(0,100)	ω	0.025	0.024	0.002	0.006	0.009	0.016	0.032	0.060	0.150	1.913	
		ω_{trim}	0.033	0.032	0.003	0.008	0.012	0.020	0.042	0.078	0.183	1.869	
	6	$\Gamma^{-1}(0.001, 0.001)$	ω	0.085	0.079	0.005	0.017	0.026	0.053	0.118	0.210	0.393	1.402
			ω_{trim}	0.094	0.089	0.005	0.018	0.028	0.059	0.133	0.237	0.432	1.376
HC(0,1)		ω	0.056	0.044	0.006	0.017	0.024	0.040	0.073	0.123	0.263	1.551	
		ω_{trim}	0.061	0.049	0.006	0.018	0.026	0.043	0.080	0.136	0.284	1.565	
HN(0,1)		ω	0.055	0.041	0.008	0.019	0.025	0.040	0.070	0.117	0.244	1.577	
		ω_{trim}	0.059	0.046	0.008	0.019	0.026	0.042	0.076	0.128	0.271	1.605	
U(0,100)		ω	0.038	0.035	0.003	0.010	0.015	0.026	0.049	0.089	0.216	1.810	
		ω_{trim}	0.044	0.040	0.003	0.011	0.017	0.029	0.057	0.103	0.244	1.803	

Tab. B.5: Deskriptive Statistiken der Gewichtungsfaktoren ω : Szenario B, $OR = 0.7$.



(a) Szenario A: $OR = 0.5$.



(b) Szenario B: $OR = 0.7$.

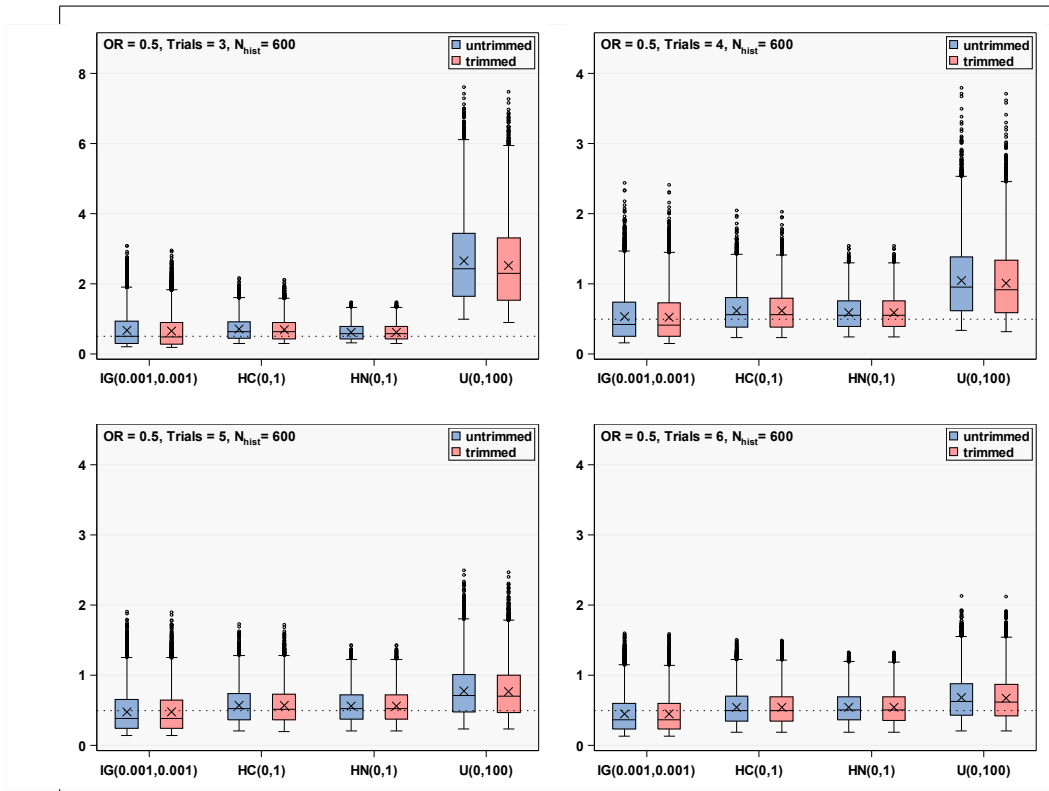
Abb. B.4: Boxplots der a-posteriori prädiktiven Erwartungswerte von θ^* .

Number of historical trials	Prior distribution for τ		Standard	10%	25%	75%	90%	Maximum	Skewness			
			Mean	Deviation	Quantil	Quantil	Quantil			Quantil		
3	$\Gamma^{-1}(0.001, 0.001)$	$\hat{\theta}^*$	-0.699	0.336	-1.985	-1.133	-0.922	-0.695	-0.468	-0.271	0.569	-0.046
		$\hat{\theta}_{trim}^*$	-0.698	0.336	-1.984	-1.134	-0.923	-0.695	-0.467	-0.272	0.575	-0.046
	HC(0,1)	$\hat{\theta}^*$	-0.699	0.337	-1.992	-1.137	-0.926	-0.694	-0.469	-0.271	0.572	-0.047
		$\hat{\theta}_{trim}^*$	-0.699	0.337	-1.988	-1.137	-0.925	-0.694	-0.469	-0.270	0.572	-0.047
	HN(0,1)	$\hat{\theta}^*$	-0.700	0.337	-1.981	-1.137	-0.924	-0.696	-0.469	-0.271	0.579	-0.046
		$\hat{\theta}_{trim}^*$	-0.700	0.337	-1.981	-1.136	-0.924	-0.696	-0.470	-0.271	0.579	-0.046
	U(0,100)	$\hat{\theta}^*$	-0.700	0.343	-2.008	-1.146	-0.929	-0.697	-0.465	-0.264	0.640	-0.045
		$\hat{\theta}_{trim}^*$	-0.700	0.341	-2.012	-1.144	-0.927	-0.697	-0.466	-0.265	0.620	-0.045
4	$\Gamma^{-1}(0.001, 0.001)$	$\hat{\theta}^*$	-0.697	0.307	-1.841	-1.087	-0.906	-0.696	-0.493	-0.310	0.552	0.006
		$\hat{\theta}_{trim}^*$	-0.697	0.307	-1.841	-1.087	-0.906	-0.696	-0.493	-0.310	0.551	0.006
	HC(0,1)	$\hat{\theta}^*$	-0.699	0.307	-1.853	-1.090	-0.909	-0.698	-0.494	-0.310	0.555	0.003
		$\hat{\theta}_{trim}^*$	-0.699	0.307	-1.852	-1.090	-0.909	-0.697	-0.494	-0.310	0.555	0.003
	HN(0,1)	$\hat{\theta}^*$	-0.699	0.307	-1.846	-1.090	-0.909	-0.699	-0.494	-0.311	0.554	0.005
		$\hat{\theta}_{trim}^*$	-0.699	0.307	-1.845	-1.089	-0.910	-0.699	-0.495	-0.311	0.554	0.005
	U(0,100)	$\hat{\theta}^*$	-0.701	0.309	-1.869	-1.093	-0.913	-0.700	-0.495	-0.311	0.553	0.001
		$\hat{\theta}_{trim}^*$	-0.701	0.309	-1.864	-1.092	-0.912	-0.700	-0.494	-0.310	0.555	0.001
5	$\Gamma^{-1}(0.001, 0.001)$	$\hat{\theta}^*$	-0.696	0.278	-1.864	-1.055	-0.883	-0.695	-0.511	-0.336	0.310	-0.036
		$\hat{\theta}_{trim}^*$	-0.696	0.278	-1.864	-1.055	-0.883	-0.695	-0.511	-0.337	0.310	-0.036
	HC(0,1)	$\hat{\theta}^*$	-0.698	0.279	-1.860	-1.058	-0.886	-0.697	-0.511	-0.339	0.303	-0.038
		$\hat{\theta}_{trim}^*$	-0.698	0.279	-1.859	-1.058	-0.885	-0.697	-0.510	-0.339	0.304	-0.038
	HN(0,1)	$\hat{\theta}^*$	-0.698	0.279	-1.864	-1.059	-0.886	-0.697	-0.511	-0.338	0.307	-0.038
		$\hat{\theta}_{trim}^*$	-0.698	0.279	-1.863	-1.058	-0.886	-0.697	-0.511	-0.338	0.307	-0.038
	U(0,100)	$\hat{\theta}^*$	-0.700	0.280	-1.872	-1.061	-0.888	-0.699	-0.513	-0.337	0.302	-0.038
		$\hat{\theta}_{trim}^*$	-0.700	0.280	-1.870	-1.061	-0.888	-0.699	-0.513	-0.338	0.303	-0.037
6	$\Gamma^{-1}(0.001, 0.001)$	$\hat{\theta}^*$	-0.696	0.266	-1.737	-1.038	-0.879	-0.696	-0.519	-0.355	0.292	0.018
		$\hat{\theta}_{trim}^*$	-0.696	0.266	-1.737	-1.037	-0.879	-0.697	-0.519	-0.355	0.292	0.018
	HC(0,1)	$\hat{\theta}^*$	-0.699	0.268	-1.751	-1.041	-0.882	-0.700	-0.521	-0.357	0.296	0.016
		$\hat{\theta}_{trim}^*$	-0.699	0.268	-1.751	-1.041	-0.882	-0.700	-0.521	-0.357	0.297	0.016
	HN(0,1)	$\hat{\theta}^*$	-0.699	0.268	-1.752	-1.041	-0.882	-0.701	-0.521	-0.355	0.296	0.018
		$\hat{\theta}_{trim}^*$	-0.699	0.268	-1.751	-1.042	-0.882	-0.700	-0.521	-0.355	0.296	0.018
	U(0,100)	$\hat{\theta}^*$	-0.701	0.268	-1.752	-1.044	-0.885	-0.701	-0.522	-0.358	0.293	0.016
		$\hat{\theta}_{trim}^*$	-0.701	0.268	-1.752	-1.044	-0.886	-0.701	-0.522	-0.358	0.292	0.016

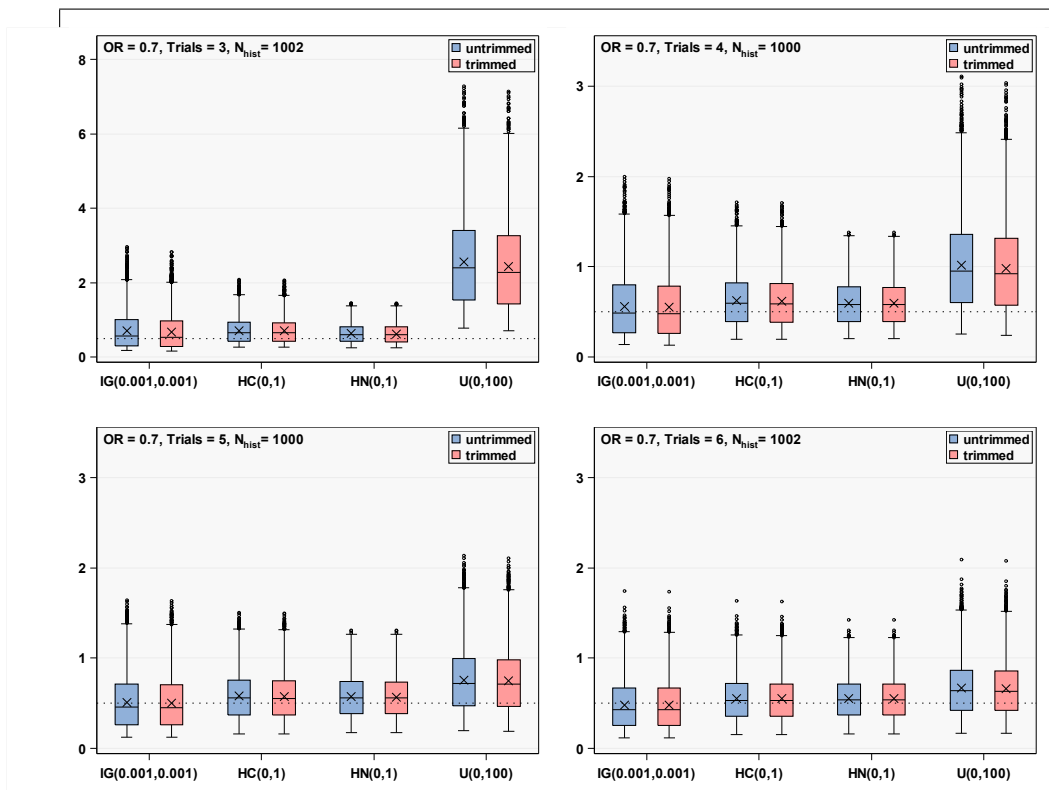
Tab. B.6: Deskriptive Statistiken der a-posteriori prädiktiven Erwartungswerte für θ^* : Szenario A, $OR = 0.5$.

Number of historical trials	Prior distribution for τ		Mean	Standard Deviation	Minimum	10% Quantil	25% Quantil	Median	75% Quantil	90% Quantil	Maximum	Skewness
3	$\Gamma^{-1}(0.001, 0.001)$	θ^*	-0.363	0.318	-1.600	-0.771	-0.576	-0.368	-0.145	0.050	0.785	0.021
		θ^*_{trim}	-0.363	0.318	-1.599	-0.769	-0.578	-0.367	-0.146	0.050	0.780	0.022
	HC(0,1)	θ^*	-0.363	0.318	-1.608	-0.769	-0.578	-0.367	-0.146	0.050	0.781	0.021
		θ^*_{trim}	-0.363	0.318	-1.606	-0.769	-0.578	-0.367	-0.145	0.050	0.782	0.021
	HN(0,1)	θ^*	-0.363	0.318	-1.605	-0.770	-0.577	-0.368	-0.147	0.050	0.774	0.021
		θ^*_{trim}	-0.363	0.318	-1.605	-0.769	-0.578	-0.367	-0.147	0.049	0.774	0.021
U(0,100)	θ^*	-0.362	0.323	-1.698	-0.770	-0.584	-0.366	-0.141	0.055	0.842	0.030	
	θ^*_{trim}	-0.362	0.322	-1.685	-0.770	-0.581	-0.366	-0.143	0.054	0.832	0.023	
4	$\Gamma^{-1}(0.001, 0.001)$	θ^*	-0.355	0.283	-1.456	-0.723	-0.549	-0.354	-0.157	0.007	0.710	-0.022
		θ^*_{trim}	-0.355	0.283	-1.453	-0.723	-0.550	-0.354	-0.156	0.007	0.710	-0.022
	HC(0,1)	θ^*	-0.355	0.283	-1.460	-0.724	-0.550	-0.355	-0.157	0.006	0.713	-0.023
		θ^*_{trim}	-0.355	0.283	-1.462	-0.724	-0.550	-0.355	-0.157	0.006	0.713	-0.023
	HN(0,1)	θ^*	-0.356	0.283	-1.451	-0.725	-0.550	-0.355	-0.157	0.006	0.707	-0.024
		θ^*_{trim}	-0.356	0.283	-1.451	-0.725	-0.550	-0.355	-0.157	0.006	0.708	-0.023
U(0,100)	θ^*	-0.356	0.284	-1.449	-0.725	-0.553	-0.355	-0.159	0.007	0.721	-0.019	
	θ^*_{trim}	-0.356	0.284	-1.447	-0.726	-0.552	-0.354	-0.158	0.006	0.724	-0.020	
5	$\Gamma^{-1}(0.001, 0.001)$	θ^*	-0.354	0.259	-1.280	-0.692	-0.528	-0.351	-0.179	-0.028	0.553	-0.038
		θ^*_{trim}	-0.354	0.259	-1.279	-0.693	-0.529	-0.351	-0.179	-0.028	0.552	-0.038
	HC(0,1)	θ^*	-0.355	0.260	-1.292	-0.693	-0.528	-0.352	-0.178	-0.026	0.559	-0.038
		θ^*_{trim}	-0.355	0.260	-1.292	-0.694	-0.528	-0.352	-0.179	-0.026	0.560	-0.038
	HN(0,1)	θ^*	-0.355	0.260	-1.298	-0.696	-0.530	-0.351	-0.179	-0.028	0.560	-0.039
		θ^*_{trim}	-0.355	0.260	-1.297	-0.696	-0.530	-0.351	-0.179	-0.028	0.560	-0.039
U(0,100)	θ^*	-0.356	0.261	-1.279	-0.695	-0.530	-0.353	-0.180	-0.029	0.563	-0.038	
	θ^*_{trim}	-0.356	0.261	-1.280	-0.695	-0.530	-0.353	-0.181	-0.030	0.565	-0.037	
6	$\Gamma^{-1}(0.001, 0.001)$	θ^*	-0.357	0.246	-1.216	-0.670	-0.526	-0.357	-0.190	-0.040	0.468	0.015
		θ^*_{trim}	-0.357	0.246	-1.216	-0.670	-0.526	-0.358	-0.190	-0.040	0.468	0.015
	HC(0,1)	θ^*	-0.358	0.247	-1.226	-0.674	-0.528	-0.359	-0.191	-0.041	0.475	0.014
		θ^*_{trim}	-0.358	0.247	-1.226	-0.674	-0.528	-0.359	-0.191	-0.041	0.475	0.014
	HN(0,1)	θ^*	-0.358	0.247	-1.230	-0.672	-0.527	-0.359	-0.191	-0.040	0.470	0.016
		θ^*_{trim}	-0.358	0.247	-1.230	-0.672	-0.527	-0.359	-0.191	-0.040	0.470	0.016
U(0,100)	θ^*	-0.359	0.247	-1.233	-0.674	-0.528	-0.360	-0.191	-0.040	0.474	0.014	
	θ^*_{trim}	-0.359	0.247	-1.233	-0.674	-0.528	-0.360	-0.191	-0.040	0.474	0.014	

Tab. B.7: Deskriptive Statistiken der a-posteriori prädiktiven Erwartungswerte für θ^* : Szenario B, $OR = 0.7$.



(a) Szenario A: $OR = 0.5$.



(b) Szenario B: $OR = 0.7$.

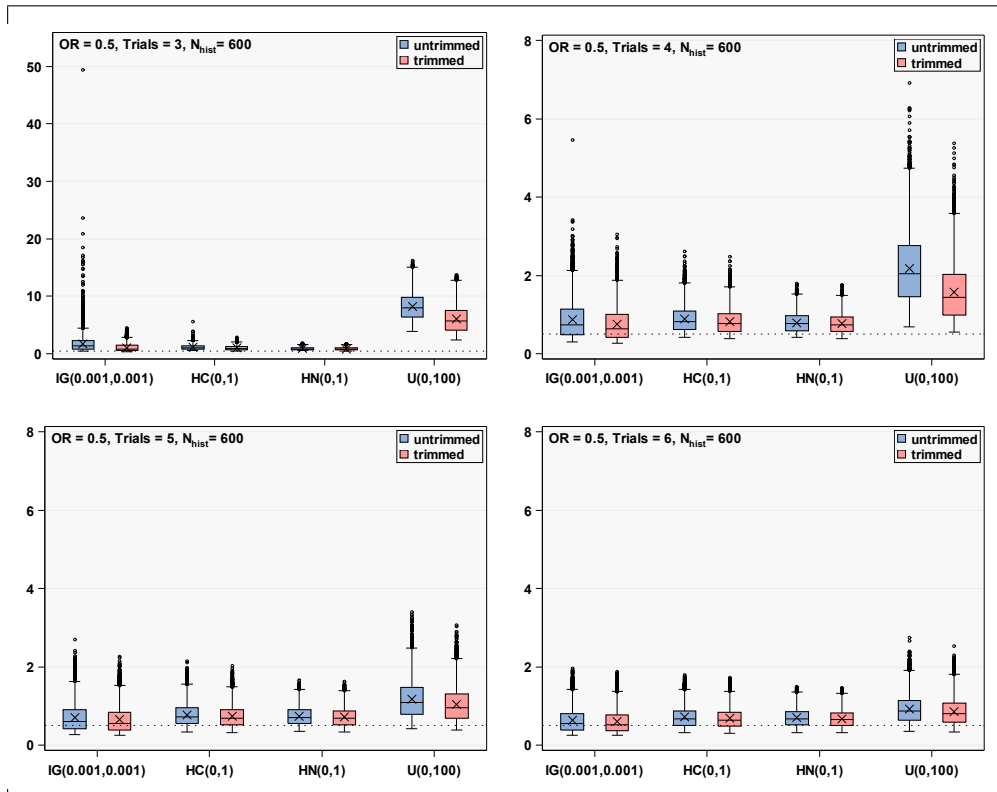
Abb. B.5: Boxplots der a-posteriori Erwartungswerte der Interstudienstandardabweichung τ .

Number of historical trials	Prior distribution for τ		Mean	Standard Deviation	Minimum	10% Quantil	25% Quantil	Median	75% Quantil	90% Quantil	Maximum	Skewness
3	$\Gamma^{-1}(0.001, 0.001)$	τ	0.68	0.48	0.20	0.24	0.30	0.51	0.94	1.40	3.09	1.27
		τ_{trim}	0.65	0.46	0.19	0.23	0.29	0.49	0.90	1.35	2.96	1.27
	HC(0,1)	τ	0.71	0.32	0.31	0.37	0.45	0.64	0.91	1.17	2.17	0.92
		τ_{trim}	0.70	0.31	0.30	0.36	0.44	0.63	0.90	1.16	2.12	0.92
	HN(0,1)	τ	0.63	0.23	0.31	0.36	0.43	0.58	0.79	0.96	1.47	0.64
		τ_{trim}	0.62	0.23	0.31	0.35	0.42	0.58	0.79	0.96	1.47	0.64
	U(0,100)	τ	2.65	1.18	1.00	1.29	1.65	2.43	3.44	4.34	7.60	0.76
		τ_{trim}	2.53	1.17	0.91	1.18	1.54	2.30	3.31	4.21	7.46	0.77
4	$\Gamma^{-1}(0.001, 0.001)$	τ	0.54	0.34	0.16	0.20	0.26	0.42	0.74	1.05	2.44	1.17
		τ_{trim}	0.53	0.34	0.15	0.19	0.25	0.41	0.73	1.03	2.41	1.17
	HC(0,1)	τ	0.62	0.28	0.23	0.31	0.39	0.56	0.80	1.01	2.04	0.83
		τ_{trim}	0.61	0.28	0.23	0.30	0.38	0.56	0.80	1.00	2.03	0.83
	HN(0,1)	τ	0.59	0.23	0.24	0.31	0.39	0.55	0.76	0.92	1.54	0.58
		τ_{trim}	0.59	0.23	0.24	0.31	0.39	0.55	0.75	0.92	1.54	0.58
	U(0,100)	τ	1.05	0.52	0.33	0.46	0.62	0.95	1.38	1.77	3.79	0.85
		τ_{trim}	1.01	0.51	0.32	0.44	0.59	0.91	1.34	1.72	3.71	0.86
5	$\Gamma^{-1}(0.001, 0.001)$	τ	0.48	0.30	0.14	0.19	0.25	0.39	0.65	0.93	1.90	1.13
		τ_{trim}	0.48	0.29	0.14	0.19	0.24	0.38	0.64	0.92	1.89	1.14
	HC(0,1)	τ	0.57	0.25	0.20	0.29	0.37	0.52	0.74	0.94	1.73	0.80
		τ_{trim}	0.57	0.25	0.20	0.28	0.36	0.52	0.73	0.94	1.72	0.80
	HN(0,1)	τ	0.56	0.23	0.21	0.29	0.38	0.53	0.72	0.89	1.43	0.60
		τ_{trim}	0.56	0.23	0.21	0.29	0.37	0.52	0.72	0.89	1.43	0.60
	U(0,100)	τ	0.78	0.37	0.23	0.36	0.48	0.71	1.01	1.30	2.49	0.81
		τ_{trim}	0.77	0.36	0.23	0.35	0.47	0.70	1.00	1.29	2.46	0.81
6	$\Gamma^{-1}(0.001, 0.001)$	τ	0.45	0.27	0.14	0.18	0.24	0.37	0.60	0.86	1.60	1.14
		τ_{trim}	0.45	0.27	0.13	0.18	0.23	0.36	0.60	0.85	1.59	1.14
	HC(0,1)	τ	0.54	0.24	0.18	0.27	0.35	0.50	0.70	0.89	1.50	0.79
		τ_{trim}	0.54	0.24	0.18	0.27	0.35	0.50	0.69	0.88	1.49	0.79
	HN(0,1)	τ	0.54	0.22	0.19	0.28	0.36	0.51	0.69	0.86	1.33	0.62
		τ_{trim}	0.54	0.22	0.19	0.28	0.36	0.51	0.69	0.86	1.33	0.62
	U(0,100)	τ	0.68	0.31	0.21	0.32	0.43	0.62	0.88	1.12	2.13	0.78
		τ_{trim}	0.67	0.31	0.21	0.32	0.42	0.62	0.87	1.11	2.12	0.78

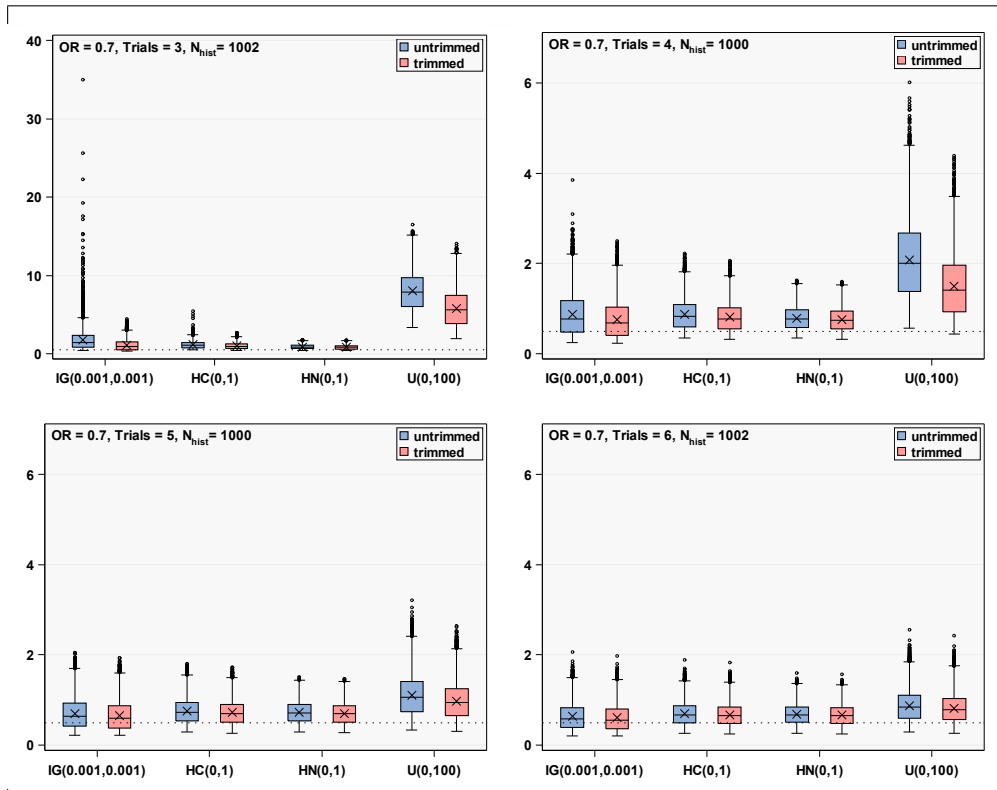
Tab. B.8: Deskriptive Statistiken für $\mathbb{E}(\tau|y)$: Szenario A, $OR = 0.5$.

Number of historical trials	Prior distribution for τ		Standard	10%	25%	75%	90%	Maximum	Skewness				
			Mean	Deviation	Quantil	Quantil	Quantil			Quantil			
3	$\Gamma^{-1}(0.001, 0.001)$	τ	0.70	0.48	0.17	0.21	0.30	0.56	1.01	1.40	2.96	1.01	
		τ_{trim}	0.67	0.46	0.17	0.20	0.28	0.54	0.97	1.34	2.82	1.01	
	HC(0,1)	τ	0.71	0.32	0.27	0.33	0.43	0.66	0.93	1.16	2.08	0.70	
		τ_{trim}	0.70	0.32	0.26	0.32	0.43	0.65	0.92	1.15	2.06	0.70	
	HN(0,1)	τ	0.63	0.24	0.26	0.32	0.42	0.61	0.81	0.96	1.45	0.42	
		τ_{trim}	0.63	0.24	0.25	0.32	0.41	0.60	0.81	0.96	1.45	0.42	
	U(0,100)	τ	2.56	1.20	0.78	1.12	1.54	2.40	3.40	4.24	7.29	0.61	
		τ_{trim}	2.44	1.18	0.71	1.01	1.43	2.27	3.27	4.10	7.14	0.62	
	4	$\Gamma^{-1}(0.001, 0.001)$	τ	0.56	0.34	0.14	0.18	0.27	0.49	0.80	1.05	2.00	0.80
			τ_{trim}	0.55	0.34	0.13	0.18	0.26	0.48	0.79	1.03	1.98	0.80
HC(0,1)		τ	0.62	0.28	0.20	0.28	0.39	0.59	0.82	1.00	1.72	0.53	
		τ_{trim}	0.62	0.28	0.19	0.28	0.39	0.59	0.81	1.00	1.70	0.53	
HN(0,1)		τ	0.60	0.24	0.20	0.29	0.40	0.58	0.77	0.92	1.38	0.31	
		τ_{trim}	0.59	0.24	0.20	0.29	0.39	0.58	0.77	0.91	1.38	0.31	
U(0,100)		τ	1.02	0.50	0.26	0.40	0.60	0.95	1.36	1.70	3.11	0.63	
		τ_{trim}	0.98	0.49	0.24	0.39	0.58	0.92	1.31	1.65	3.04	0.64	
5		$\Gamma^{-1}(0.001, 0.001)$	τ	0.51	0.29	0.12	0.18	0.26	0.45	0.71	0.93	1.64	0.71
			τ_{trim}	0.51	0.29	0.12	0.18	0.26	0.45	0.70	0.92	1.63	0.71
	HC(0,1)	τ	0.58	0.25	0.16	0.27	0.37	0.56	0.75	0.92	1.51	0.47	
		τ_{trim}	0.58	0.25	0.16	0.26	0.37	0.55	0.75	0.92	1.50	0.47	
	HN(0,1)	τ	0.57	0.23	0.18	0.27	0.38	0.56	0.74	0.88	1.31	0.30	
		τ_{trim}	0.57	0.23	0.17	0.27	0.38	0.56	0.74	0.88	1.31	0.30	
	U(0,100)	τ	0.76	0.35	0.19	0.32	0.47	0.72	0.99	1.24	2.13	0.54	
		τ_{trim}	0.75	0.34	0.19	0.32	0.46	0.71	0.98	1.23	2.11	0.54	
	6	$\Gamma^{-1}(0.001, 0.001)$	τ	0.48	0.27	0.12	0.18	0.25	0.43	0.67	0.86	1.75	0.69
			τ_{trim}	0.48	0.26	0.12	0.17	0.25	0.43	0.67	0.86	1.74	0.69
HC(0,1)		τ	0.55	0.23	0.16	0.25	0.36	0.53	0.72	0.88	1.63	0.46	
		τ_{trim}	0.55	0.23	0.15	0.25	0.36	0.53	0.71	0.87	1.63	0.46	
HN(0,1)		τ	0.55	0.22	0.16	0.26	0.37	0.54	0.71	0.85	1.42	0.31	
		τ_{trim}	0.55	0.22	0.16	0.26	0.37	0.54	0.71	0.85	1.42	0.31	
U(0,100)		τ	0.67	0.30	0.17	0.29	0.42	0.64	0.87	1.07	2.09	0.51	
		τ_{trim}	0.66	0.29	0.16	0.29	0.42	0.63	0.86	1.07	2.08	0.51	

Tab. B.9: Deskriptive Statistiken für $\mathbb{E}(\tau|y)$: Szenario B, $OR = 0.7$.



(a) Szenario A: $OR = 0.5$.



(b) Szenario B: $OR = 0.7$.

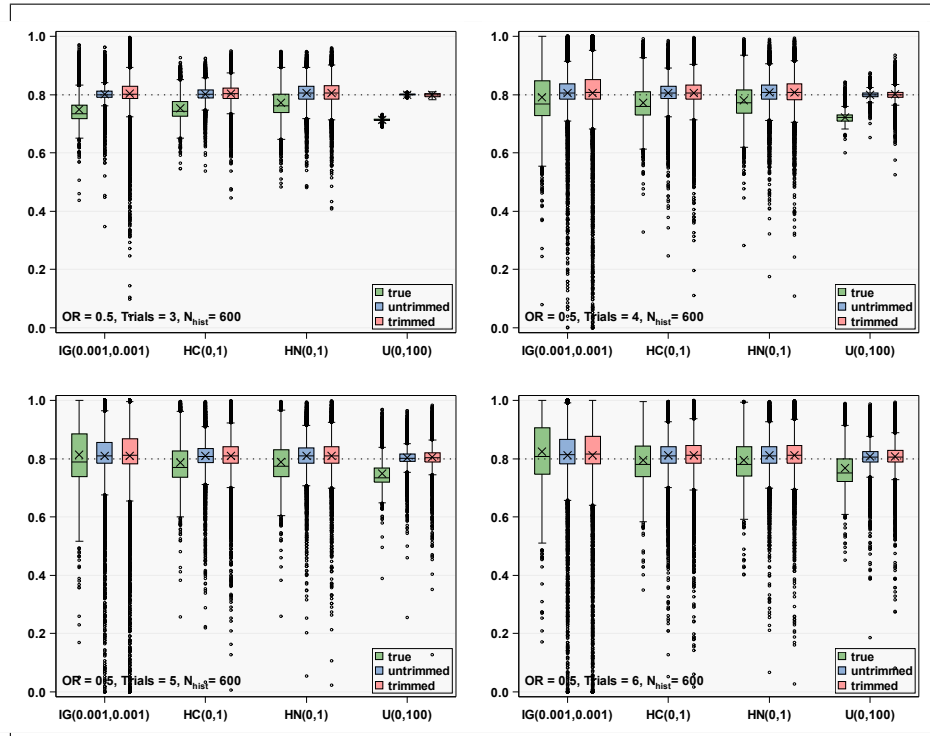
Abb. B.6: Boxplots der a-posteriori prädiktiven Standardabweichung $V_\tau^{1/2}$ von θ^* .

Number of historical trials	Prior distribution for τ		Standard		10%	25%	75%	90%	Maximum	Skewness			
			Mean	Deviation	Minimum	Quantil	Quantil	Median			Quantil	Quantil	
3	$\Gamma^1(0.001, 0.001)$	$V_{\tau}^{1/2}$	1.78	1.44	0.45	0.64	0.84	1.39	2.30	3.34	49.41	6.59	
		$V_{\tau}^{1/2 \text{ trim}}$	1.09	0.68	0.36	0.44	0.54	0.86	1.47	2.09	4.48	1.23	
	HC(0,1)	$V_{\tau}^{1/2}$	1.17	0.43	0.59	0.71	0.82	1.08	1.43	1.77	5.61	1.07	
		$V_{\tau}^{1/2 \text{ trim}}$	1.02	0.38	0.52	0.60	0.71	0.94	1.26	1.56	2.81	0.92	
	HN(0,1)	$V_{\tau}^{1/2}$	0.89	0.24	0.54	0.60	0.68	0.85	1.05	1.23	1.79	0.62	
		$V_{\tau}^{1/2 \text{ trim}}$	0.85	0.24	0.50	0.56	0.64	0.81	1.02	1.20	1.76	0.62	
	U(0,100)	$V_{\tau}^{1/2}$	8.23	2.24	3.89	5.49	6.34	8.01	9.83	11.39	16.15	0.47	
		$V_{\tau}^{1/2 \text{ trim}}$	6.01	2.21	2.42	3.33	4.14	5.73	7.58	9.14	13.75	0.55	
	4	$\Gamma^1(0.001, 0.001)$	$V_{\tau}^{1/2}$	0.87	0.47	0.30	0.39	0.49	0.73	1.15	1.54	5.46	1.20
			$V_{\tau}^{1/2 \text{ trim}}$	0.76	0.41	0.28	0.34	0.42	0.64	1.01	1.36	3.04	1.12
		HC(0,1)	$V_{\tau}^{1/2}$	0.89	0.32	0.41	0.52	0.62	0.83	1.10	1.34	2.62	0.83
			$V_{\tau}^{1/2 \text{ trim}}$	0.83	0.31	0.38	0.47	0.57	0.77	1.03	1.26	2.48	0.83
HN(0,1)		$V_{\tau}^{1/2}$	0.80	0.24	0.42	0.50	0.59	0.76	0.97	1.13	1.79	0.55	
		$V_{\tau}^{1/2 \text{ trim}}$	0.77	0.24	0.39	0.48	0.57	0.74	0.94	1.11	1.75	0.55	
U(0,100)		$V_{\tau}^{1/2}$	2.18	0.90	0.69	1.11	1.46	2.04	2.77	3.42	6.91	0.70	
		$V_{\tau}^{1/2 \text{ trim}}$	1.57	0.71	0.56	0.76	0.99	1.44	2.03	2.55	5.37	0.85	
5		$\Gamma^1(0.001, 0.001)$	$V_{\tau}^{1/2}$	0.70	0.35	0.27	0.34	0.42	0.60	0.91	1.22	2.69	1.08
			$V_{\tau}^{1/2 \text{ trim}}$	0.65	0.33	0.25	0.32	0.39	0.56	0.85	1.15	2.27	1.08
		HC(0,1)	$V_{\tau}^{1/2}$	0.78	0.28	0.34	0.45	0.55	0.72	0.95	1.18	2.15	0.81
			$V_{\tau}^{1/2 \text{ trim}}$	0.74	0.27	0.33	0.43	0.52	0.69	0.91	1.13	2.02	0.81
	HN(0,1)	$V_{\tau}^{1/2}$	0.74	0.23	0.35	0.46	0.55	0.71	0.90	1.07	1.66	0.57	
		$V_{\tau}^{1/2 \text{ trim}}$	0.72	0.23	0.33	0.44	0.53	0.68	0.87	1.05	1.63	0.58	
	U(0,100)	$V_{\tau}^{1/2}$	1.17	0.48	0.41	0.62	0.79	1.09	1.47	1.84	3.39	0.81	
		$V_{\tau}^{1/2 \text{ trim}}$	1.03	0.43	0.38	0.54	0.69	0.95	1.30	1.64	3.06	0.82	
	6	$\Gamma^1(0.001, 0.001)$	$V_{\tau}^{1/2}$	0.63	0.30	0.26	0.32	0.39	0.55	0.81	1.07	1.95	1.06
			$V_{\tau}^{1/2 \text{ trim}}$	0.60	0.29	0.24	0.30	0.37	0.52	0.77	1.03	1.87	1.08
		HC(0,1)	$V_{\tau}^{1/2}$	0.71	0.25	0.32	0.42	0.51	0.67	0.88	1.07	1.79	0.79
			$V_{\tau}^{1/2 \text{ trim}}$	0.69	0.25	0.30	0.40	0.49	0.64	0.84	1.04	1.73	0.79
HN(0,1)		$V_{\tau}^{1/2}$	0.70	0.22	0.33	0.43	0.52	0.67	0.85	1.02	1.49	0.58	
		$V_{\tau}^{1/2 \text{ trim}}$	0.68	0.22	0.31	0.41	0.50	0.65	0.83	1.00	1.46	0.59	
U(0,100)		$V_{\tau}^{1/2}$	0.92	0.36	0.36	0.51	0.64	0.86	1.15	1.42	2.75	0.78	
		$V_{\tau}^{1/2 \text{ trim}}$	0.86	0.34	0.33	0.47	0.59	0.81	1.08	1.34	2.52	0.78	

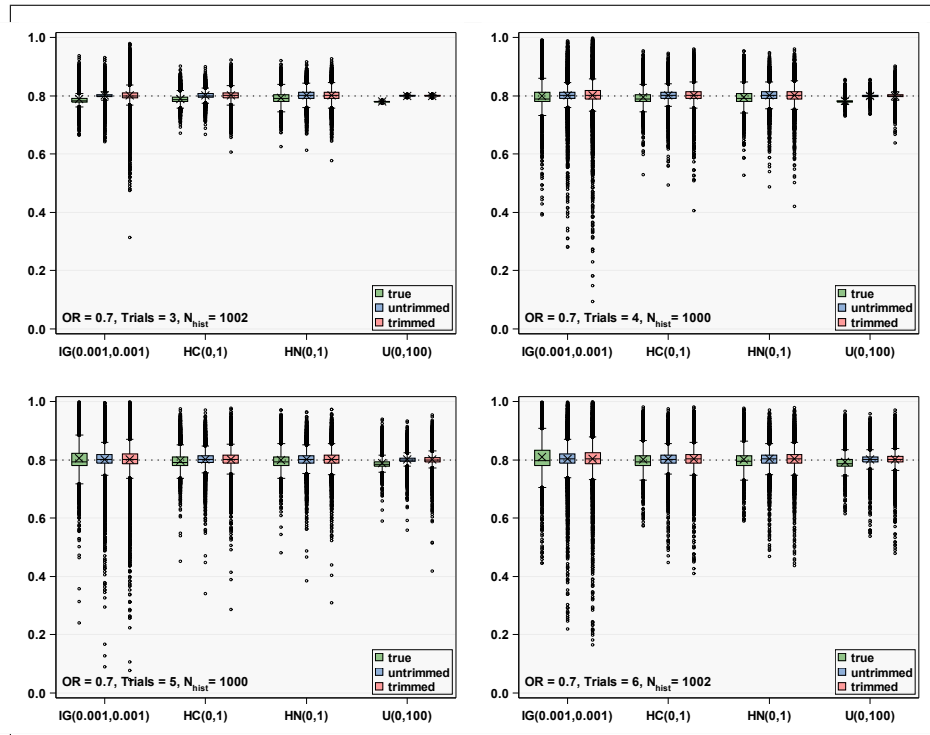
Tab. B.10: Deskriptive Statistiken für $V_{\tau}^{1/2}$: Szenario A, $OR = 0.5$.

Number of historical trials	Prior distribution for τ		Standard		10%	25%	75%	90%	Maximum	Skewness			
			Mean	Deviation	Minimum	Quantil	Quantil	Median			Quantil	Quantil	
3	$\Gamma^{-1}(0.001, 0.001)$	$V_{\tau}^{1/2}$	1.76	1.39	0.39	0.57	0.81	1.41	2.34	3.33	34.99	4.50	
		$V_{\tau}^{1/2 \text{ trim}}$	1.08	0.68	0.31	0.38	0.51	0.90	1.52	2.06	4.43	0.99	
	HC(0,1)	$V_{\tau}^{1/2}$	1.15	0.44	0.52	0.64	0.79	1.08	1.43	1.74	5.44	1.03	
		$V_{\tau}^{1/2 \text{ trim}}$	1.00	0.39	0.45	0.54	0.67	0.94	1.26	1.54	2.70	0.72	
	HN(0,1)	$V_{\tau}^{1/2}$	0.87	0.26	0.45	0.55	0.65	0.85	1.06	1.23	1.76	0.42	
		$V_{\tau}^{1/2 \text{ trim}}$	0.84	0.26	0.42	0.51	0.62	0.82	1.03	1.19	1.73	0.41	
	U(0,100)	$V_{\tau}^{1/2}$	8.01	2.32	3.34	5.01	6.07	7.86	9.73	11.18	16.51	0.32	
		$V_{\tau}^{1/2 \text{ trim}}$	5.80	2.27	1.91	2.89	3.87	5.62	7.47	8.94	14.06	0.40	
	4	$\Gamma^{-1}(0.001, 0.001)$	$V_{\tau}^{1/2}$	0.87	0.46	0.25	0.35	0.48	0.78	1.17	1.50	3.85	0.84
			$V_{\tau}^{1/2 \text{ trim}}$	0.76	0.41	0.23	0.30	0.42	0.68	1.04	1.33	2.50	0.79
		HC(0,1)	$V_{\tau}^{1/2}$	0.87	0.32	0.35	0.47	0.60	0.83	1.09	1.31	2.22	0.57
			$V_{\tau}^{1/2 \text{ trim}}$	0.81	0.31	0.32	0.43	0.55	0.78	1.02	1.23	2.06	0.56
HN(0,1)		$V_{\tau}^{1/2}$	0.79	0.25	0.35	0.46	0.58	0.77	0.97	1.12	1.63	0.29	
		$V_{\tau}^{1/2 \text{ trim}}$	0.76	0.25	0.33	0.43	0.55	0.75	0.94	1.09	1.60	0.29	
U(0,100)		$V_{\tau}^{1/2}$	2.08	0.88	0.56	0.98	1.38	2.00	2.68	3.29	6.02	0.53	
		$V_{\tau}^{1/2 \text{ trim}}$	1.50	0.69	0.44	0.66	0.94	1.41	1.96	2.44	4.39	0.65	
5		$\Gamma^{-1}(0.001, 0.001)$	$V_{\tau}^{1/2}$	0.70	0.34	0.22	0.31	0.42	0.64	0.93	1.18	2.04	0.70
			$V_{\tau}^{1/2 \text{ trim}}$	0.66	0.32	0.21	0.29	0.39	0.60	0.87	1.12	1.93	0.71
		HC(0,1)	$V_{\tau}^{1/2}$	0.76	0.27	0.29	0.41	0.54	0.73	0.94	1.14	1.80	0.50
			$V_{\tau}^{1/2 \text{ trim}}$	0.72	0.27	0.27	0.39	0.51	0.70	0.90	1.09	1.73	0.50
	HN(0,1)	$V_{\tau}^{1/2}$	0.73	0.24	0.30	0.42	0.54	0.72	0.90	1.05	1.51	0.29	
		$V_{\tau}^{1/2 \text{ trim}}$	0.70	0.23	0.28	0.40	0.51	0.69	0.87	1.02	1.48	0.29	
	U(0,100)	$V_{\tau}^{1/2}$	1.11	0.45	0.34	0.55	0.74	1.06	1.41	1.74	3.21	0.57	
		$V_{\tau}^{1/2 \text{ trim}}$	0.98	0.41	0.31	0.48	0.65	0.94	1.25	1.55	2.64	0.57	
	6	$\Gamma^{-1}(0.001, 0.001)$	$V_{\tau}^{1/2}$	0.63	0.29	0.21	0.29	0.39	0.58	0.83	1.05	2.07	0.68
			$V_{\tau}^{1/2 \text{ trim}}$	0.61	0.28	0.20	0.27	0.37	0.56	0.80	1.01	1.98	0.68
		HC(0,1)	$V_{\tau}^{1/2}$	0.70	0.25	0.26	0.38	0.50	0.67	0.87	1.04	1.89	0.49
			$V_{\tau}^{1/2 \text{ trim}}$	0.67	0.24	0.25	0.36	0.47	0.65	0.84	1.01	1.84	0.49
HN(0,1)		$V_{\tau}^{1/2}$	0.69	0.23	0.27	0.39	0.50	0.67	0.85	1.00	1.60	0.31	
		$V_{\tau}^{1/2 \text{ trim}}$	0.67	0.22	0.25	0.37	0.49	0.65	0.83	0.97	1.57	0.32	
U(0,100)		$V_{\tau}^{1/2}$	0.88	0.34	0.29	0.45	0.60	0.84	1.10	1.34	2.56	0.55	
		$V_{\tau}^{1/2 \text{ trim}}$	0.82	0.32	0.27	0.42	0.56	0.79	1.04	1.27	2.43	0.54	

Tab. B.11: Deskriptive Statistiken für $V_{\tau}^{1/2}$: Szenario B, $OR = 0.7$.



(a) Szenario A: $OR = 0.5$.



(b) Szenario B: $OR = 0.7$.

Abb. B.7: Boxplots der bayesianischen Power.

Für die Poweranalyse wurden der a-posteriori prädiktive Erwartungswert und die Varianz des jeweiligen Modells als informative a-priori Parameter für das Normal-Normal-Modell der neuen Studie benutzt. Für die „wahre“ Power wurden die Schätzer aus dem vollen Modell angewendet. Die geplante Fallzahl N_{ref} der neuen Studie wurde um die getrimmte oder ungetrimmte virtuelle Fallzahl N_{ef} bzw. N_{true} reduziert.

			Min	1% Quantil	5% Quantil	10% Quantil	25% Quantil	Median	75% Quantil	90% Quantil	95% Quantil	99% Quantil	Max	
3	$\Gamma^1(0.001,0.001)$	Power	0.347	0.696	0.757	0.778	0.791	0.800	0.811	0.836	0.856	0.899	0.963	
		Power _{trim}	0.042	0.528	0.694	0.749	0.787	0.802	0.829	0.883	0.919	0.969	0.995	
	HC(0,1)	Power	0.538	0.715	0.757	0.772	0.788	0.801	0.816	0.837	0.852	0.881	0.924	
		Power _{trim}	0.446	0.682	0.740	0.764	0.787	0.803	0.822	0.850	0.869	0.905	0.950	
	HN(0,1)	Power	0.481	0.679	0.736	0.759	0.784	0.805	0.828	0.855	0.874	0.908	0.948	
		Power _{trim}	0.407	0.659	0.727	0.755	0.784	0.806	0.831	0.861	0.881	0.918	0.960	
	U(0,100)	Power	0.790	0.791	0.793	0.794	0.799	0.801	0.802	0.804	0.804	0.806	0.808	
		Power _{trim}	0.782	0.791	0.792	0.793	0.794	0.800	0.802	0.804	0.805	0.807	0.809	
	4	$\Gamma^1(0.001,0.001)$	Power	0.001	0.439	0.675	0.743	0.785	0.805	0.837	0.898	0.938	0.986	1.000
			Power _{trim}	0.000	0.269	0.625	0.720	0.783	0.808	0.851	0.931	0.971	0.998	1.000
		HC(0,1)	Power	0.247	0.631	0.730	0.759	0.786	0.805	0.828	0.861	0.886	0.930	0.986
			Power _{trim}	0.112	0.589	0.715	0.753	0.785	0.806	0.833	0.873	0.901	0.948	0.995
HN(0,1)		Power	0.177	0.617	0.722	0.753	0.785	0.808	0.833	0.869	0.893	0.933	0.990	
		Power _{trim}	0.108	0.591	0.712	0.749	0.783	0.808	0.836	0.875	0.902	0.945	0.995	
U(0,100)		Power	0.653	0.766	0.783	0.787	0.792	0.799	0.805	0.812	0.817	0.834	0.874	
		Power _{trim}	0.526	0.731	0.770	0.782	0.791	0.800	0.808	0.823	0.837	0.866	0.934	
5		$\Gamma^1(0.001,0.001)$	Power	0.000	0.351	0.652	0.726	0.784	0.810	0.856	0.928	0.968	0.999	1.000
			Power _{trim}	0.000	0.225	0.611	0.712	0.782	0.812	0.867	0.950	0.986	1.000	1.000
		HC(0,1)	Power	0.034	0.609	0.721	0.753	0.786	0.808	0.835	0.877	0.907	0.956	0.995
			Power _{trim}	0.006	0.575	0.710	0.748	0.784	0.809	0.840	0.886	0.918	0.970	0.998
	HN(0,1)	Power	0.055	0.614	0.719	0.751	0.785	0.809	0.837	0.876	0.905	0.954	0.994	
		Power _{trim}	0.024	0.582	0.711	0.748	0.784	0.810	0.841	0.883	0.915	0.964	0.997	
	U(0,100)	Power	0.254	0.706	0.761	0.777	0.790	0.801	0.815	0.837	0.856	0.900	0.965	
		Power _{trim}	0.128	0.672	0.750	0.771	0.789	0.803	0.820	0.848	0.874	0.924	0.983	
	6	$\Gamma^1(0.001,0.001)$	Power	0.000	0.289	0.627	0.718	0.783	0.813	0.866	0.945	0.982	1.000	1.000
			Power _{trim}	0.000	0.167	0.593	0.707	0.782	0.816	0.877	0.963	0.994	1.000	1.000
		HC(0,1)	Power	0.053	0.575	0.706	0.749	0.785	0.810	0.842	0.888	0.919	0.971	0.996
			Power _{trim}	0.016	0.547	0.695	0.743	0.785	0.811	0.846	0.897	0.930	0.980	0.999
HN(0,1)		Power	0.066	0.588	0.708	0.749	0.785	0.811	0.842	0.886	0.915	0.967	0.995	
		Power _{trim}	0.027	0.563	0.701	0.746	0.785	0.812	0.845	0.893	0.924	0.975	0.998	
U(0,100)		Power	0.185	0.664	0.743	0.767	0.789	0.805	0.825	0.856	0.883	0.935	0.987	
		Power _{trim}	0.083	0.633	0.734	0.763	0.788	0.806	0.828	0.866	0.897	0.954	0.994	

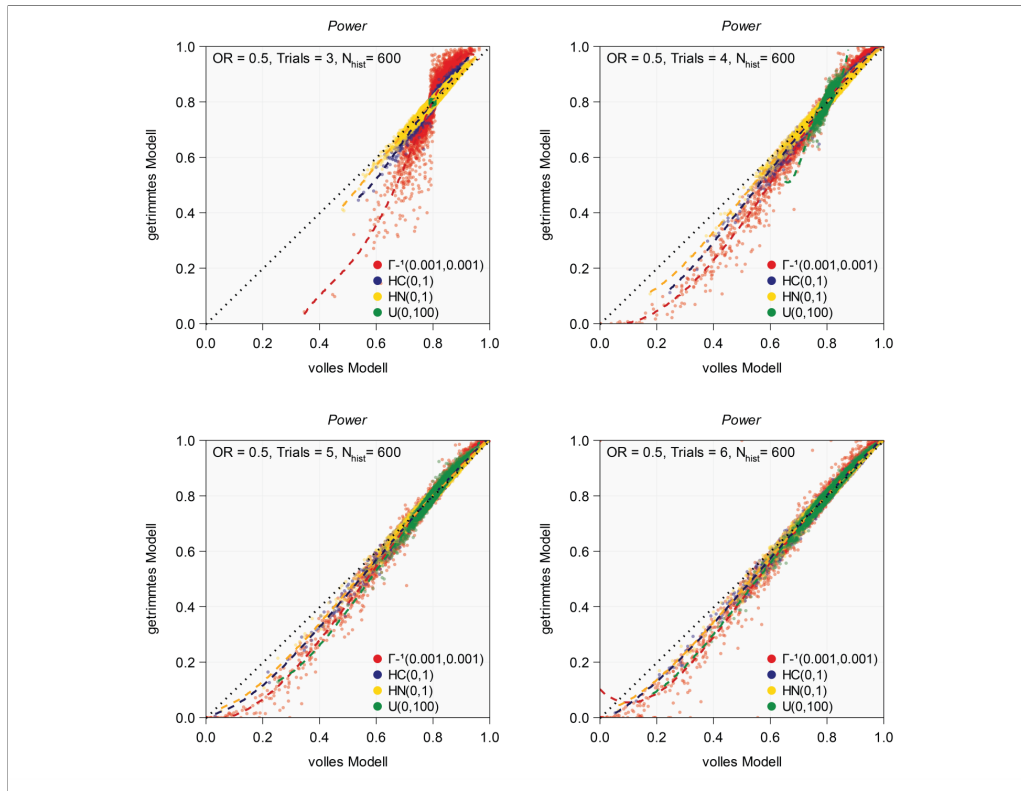
0.0 - 0.1	0.2 - 0.3	0.4 - 0.5	0.55 - 0.60	0.65 - 0.70	0.75 - 0.80	0.85 - 0.90	0.95 - 1.00
0.1 - 0.2	0.3 - 0.4	0.5 - 0.55	0.60 - 0.65	0.70 - 0.75	0.80 - 0.85	0.90 - 0.95	

Tab. B.12: Deskriptive Statistiken für die Power der neuen Studie: Szenario A, $OR = 0.5$.

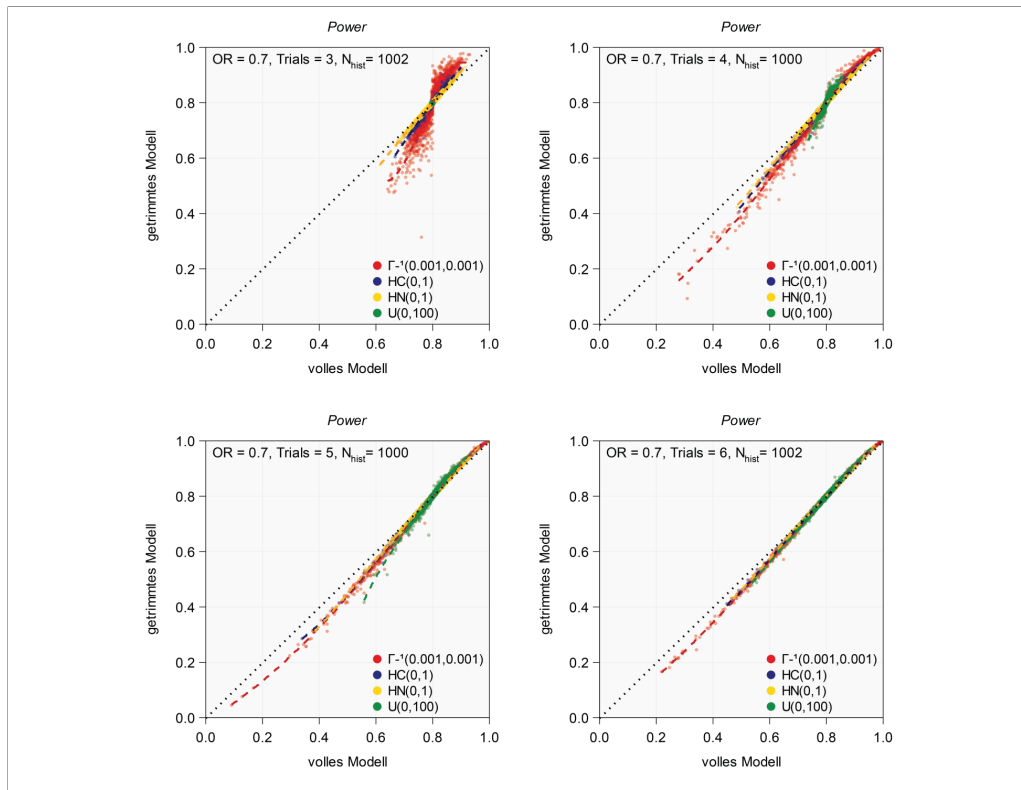
Number of historical trials	Prior distribution for τ		Min	1% Quantil	5% Quantil	10% Quantil	25% Quantil	Median	75% Quantil	90% Quantil	95% Quantil	99% Quantil	Max
3	$\Gamma^1(0.001,0.001)$	Power	0.643	0.735	0.774	0.787	0.796	0.799	0.804	0.817	0.831	0.863	0.931
		Power _{trim}	0.314	0.645	0.737	0.769	0.793	0.800	0.811	0.840	0.868	0.920	0.979
	HC(0,1)	Power	0.668	0.751	0.776	0.785	0.794	0.800	0.807	0.818	0.827	0.848	0.900
		Power _{trim}	0.606	0.732	0.767	0.780	0.793	0.800	0.809	0.824	0.837	0.864	0.923
	HN(0,1)	Power	0.613	0.733	0.765	0.778	0.791	0.801	0.812	0.826	0.838	0.864	0.917
		Power _{trim}	0.577	0.723	0.761	0.775	0.791	0.802	0.813	0.829	0.843	0.871	0.928
	U(0,100)	Power	0.794	0.796	0.797	0.798	0.799	0.800	0.801	0.801	0.802	0.803	0.805
		Power _{trim}	0.793	0.795	0.797	0.797	0.798	0.799	0.800	0.801	0.802	0.803	0.805
4	$\Gamma^1(0.001,0.001)$	Power	0.280	0.632	0.735	0.767	0.792	0.800	0.813	0.844	0.871	0.933	0.987
		Power _{trim}	0.094	0.563	0.713	0.755	0.789	0.801	0.817	0.857	0.891	0.956	0.998
	HC(0,1)	Power	0.493	0.714	0.762	0.777	0.792	0.801	0.811	0.828	0.843	0.881	0.946
		Power _{trim}	0.406	0.695	0.755	0.773	0.791	0.801	0.813	0.833	0.851	0.896	0.961
	HN(0,1)	Power	0.487	0.711	0.760	0.774	0.790	0.801	0.813	0.830	0.846	0.884	0.948
		Power _{trim}	0.421	0.697	0.755	0.772	0.790	0.802	0.814	0.833	0.851	0.894	0.961
	U(0,100)	Power	0.736	0.779	0.792	0.794	0.797	0.799	0.801	0.805	0.809	0.822	0.856
		Power _{trim}	0.639	0.758	0.783	0.790	0.796	0.799	0.803	0.811	0.820	0.846	0.902
5	$\Gamma^1(0.001,0.001)$	Power	0.090	0.592	0.722	0.761	0.789	0.802	0.818	0.854	0.887	0.947	0.996
		Power _{trim}	0.046	0.545	0.707	0.754	0.787	0.802	0.821	0.862	0.899	0.960	0.999
	HC(0,1)	Power	0.341	0.696	0.756	0.774	0.790	0.801	0.813	0.833	0.852	0.893	0.970
		Power _{trim}	0.286	0.679	0.750	0.771	0.790	0.802	0.815	0.837	0.859	0.905	0.978
	HN(0,1)	Power	0.385	0.698	0.756	0.774	0.790	0.802	0.814	0.834	0.851	0.892	0.964
		Power _{trim}	0.310	0.686	0.752	0.771	0.789	0.802	0.816	0.837	0.856	0.900	0.972
	U(0,100)	Power	0.559	0.740	0.776	0.787	0.795	0.800	0.806	0.817	0.829	0.859	0.932
		Power _{trim}	0.418	0.721	0.769	0.783	0.794	0.800	0.808	0.822	0.838	0.875	0.953
6	$\Gamma^1(0.001,0.001)$	Power	0.219	0.590	0.719	0.757	0.788	0.803	0.821	0.862	0.896	0.954	0.997
		Power _{trim}	0.166	0.557	0.705	0.751	0.787	0.803	0.823	0.869	0.906	0.964	0.999
	HC(0,1)	Power	0.448	0.689	0.753	0.772	0.790	0.802	0.816	0.839	0.860	0.904	0.974
		Power _{trim}	0.411	0.676	0.747	0.769	0.789	0.802	0.817	0.843	0.865	0.915	0.982
	HN(0,1)	Power	0.469	0.694	0.754	0.772	0.789	0.802	0.816	0.838	0.857	0.900	0.971
		Power _{trim}	0.436	0.681	0.749	0.770	0.789	0.803	0.817	0.841	0.862	0.909	0.978
	U(0,100)	Power	0.537	0.720	0.768	0.781	0.793	0.801	0.810	0.827	0.842	0.882	0.958
		Power _{trim}	0.479	0.706	0.762	0.778	0.792	0.801	0.811	0.831	0.849	0.894	0.972

0.0 - 0.1	0.2 - 0.3	0.4 - 0.5	0.55 - 0.60	0.65 - 0.70	0.75 - 0.80	0.85 - 0.90	0.95 - 1.00
0.1 - 0.2	0.3 - 0.4	0.5 - 0.55	0.60 - 0.65	0.70 - 0.75	0.80 - 0.85	0.90 - 0.95	

Tab. B.13: Deskriptive Statistiken für die Power der neuen Studie: Szenario B, $OR = 0.7$.

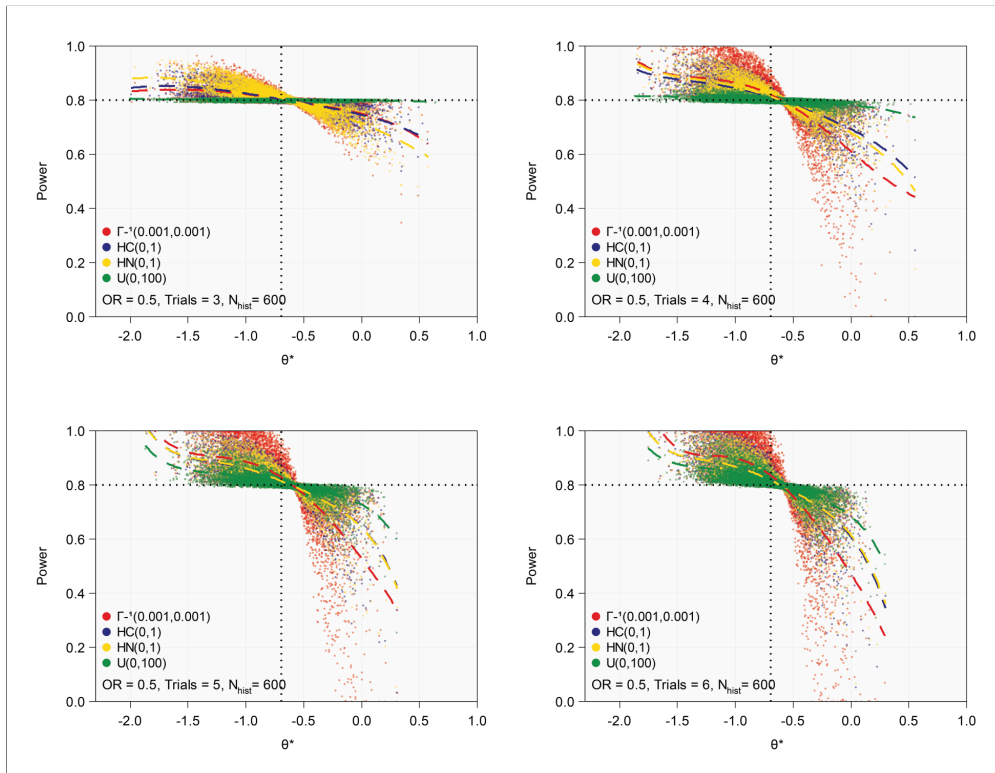


(a) Szenario A: $OR = 0.5$.

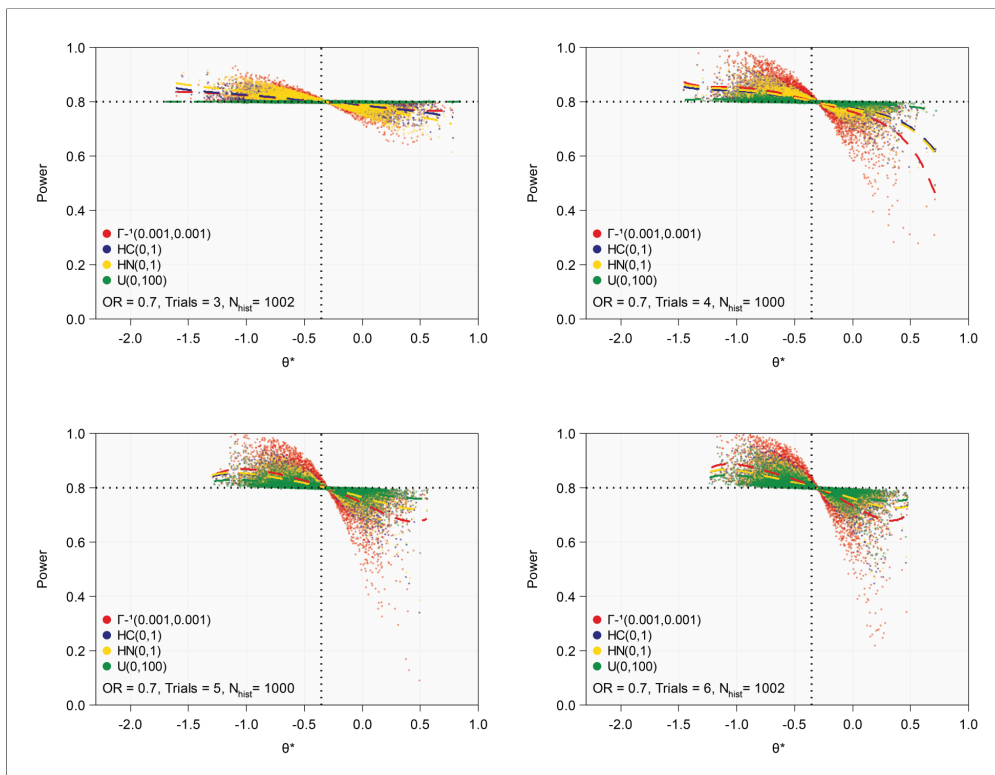


(b) Szenario B: $OR = 0.7$.

Abb. B.8: Scatterplots: Vergleich der bayesianischen Power im vollen und getrimmten Modell.

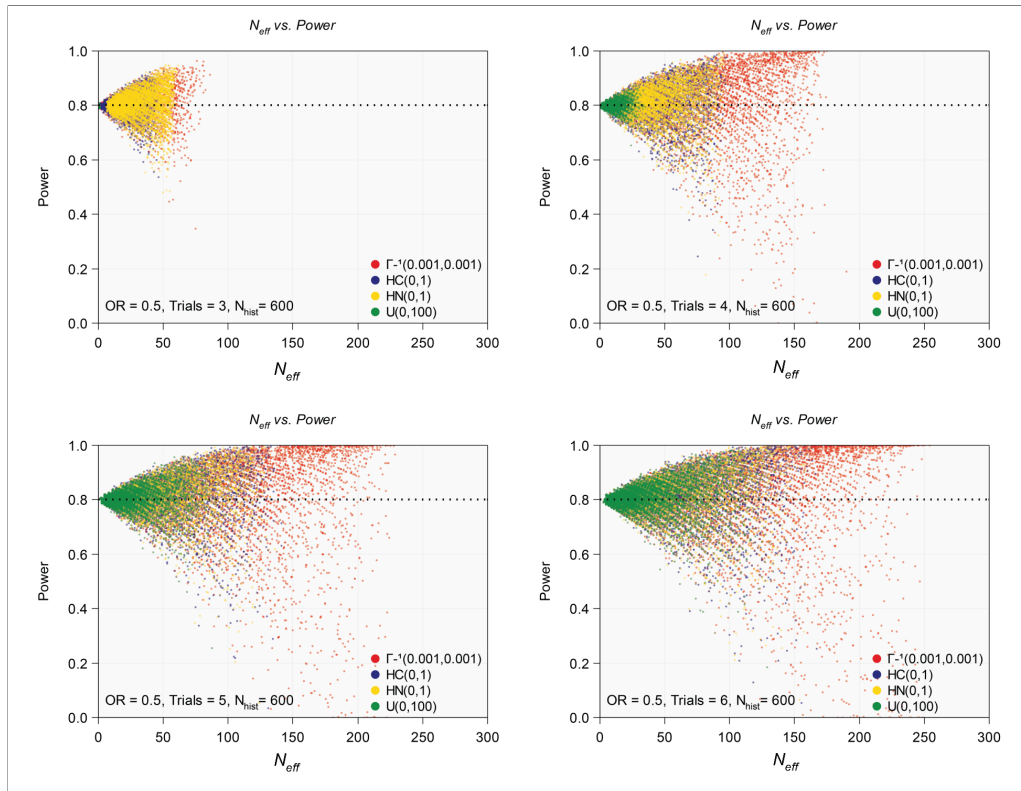


(a) Szenario A: $OR = 0.5$.

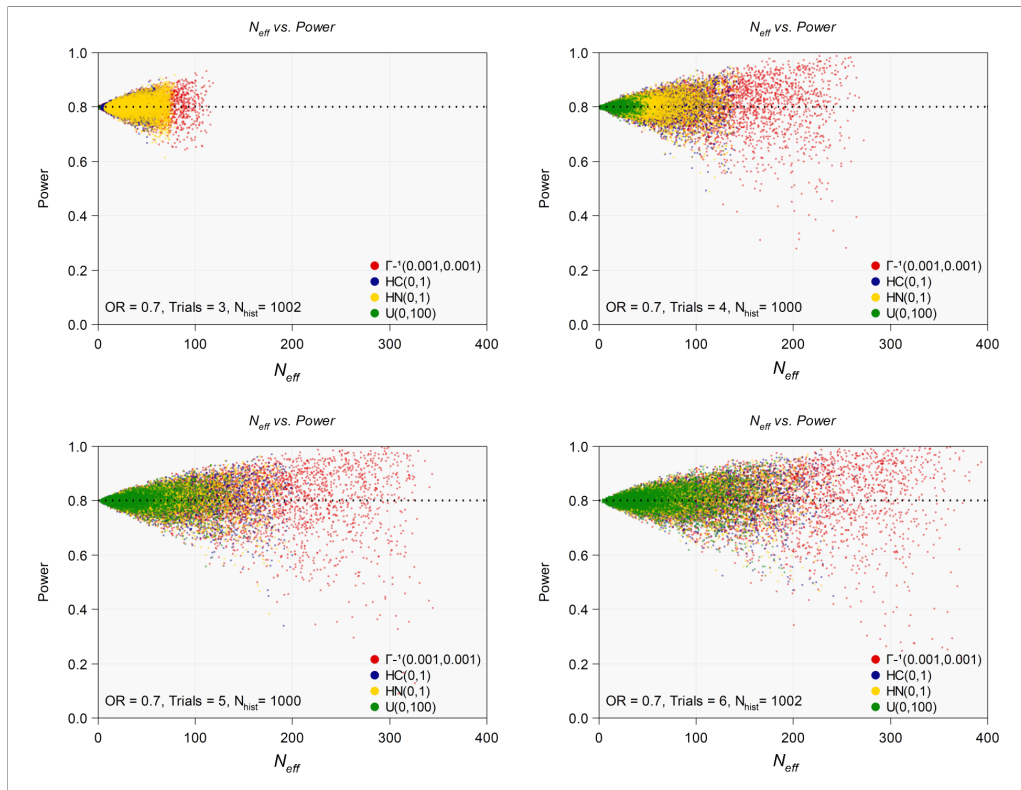


(b) Szenario B: $OR = 0.7$.

Abb. B.9: Scatterplots: Vergleich der a-posteriori prädiktiven Erwartungswerte von θ^* und der bayesianischen Power.

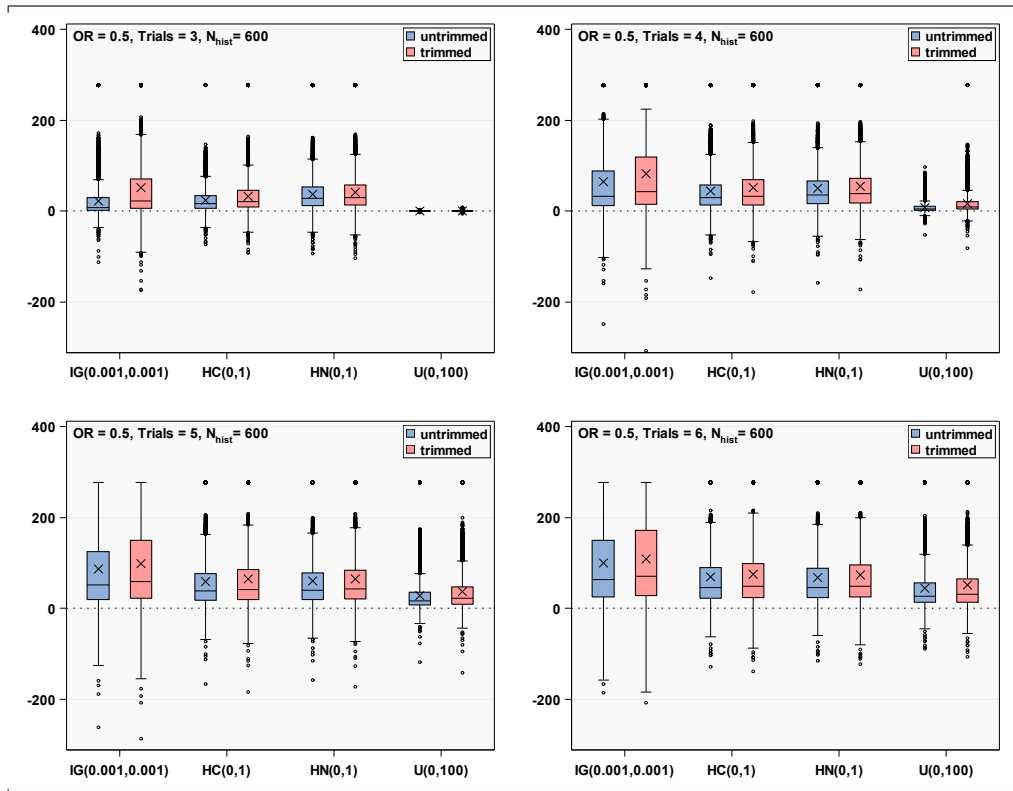


(a) Szenario A: $OR = 0.5$.

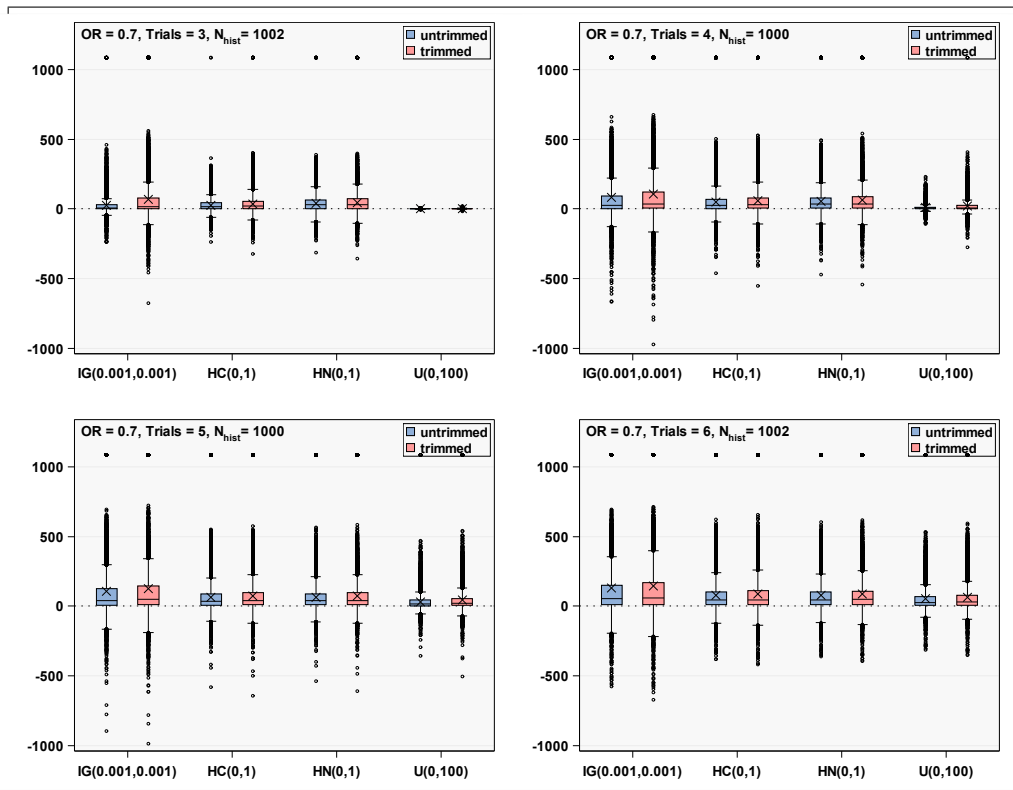


(b) Szenario B: $OR = 0.7$.

Abb. B.10: Scatterplots: Zusammenhang zwischen N_{eff} und bayesischer Power.



(a) Szenario A: $OR = 0.5$.



(b) Szenario B: $OR = 0.7$.

Abb. B.11: Boxplots der Prior Effective Sample Size N_{prior} .

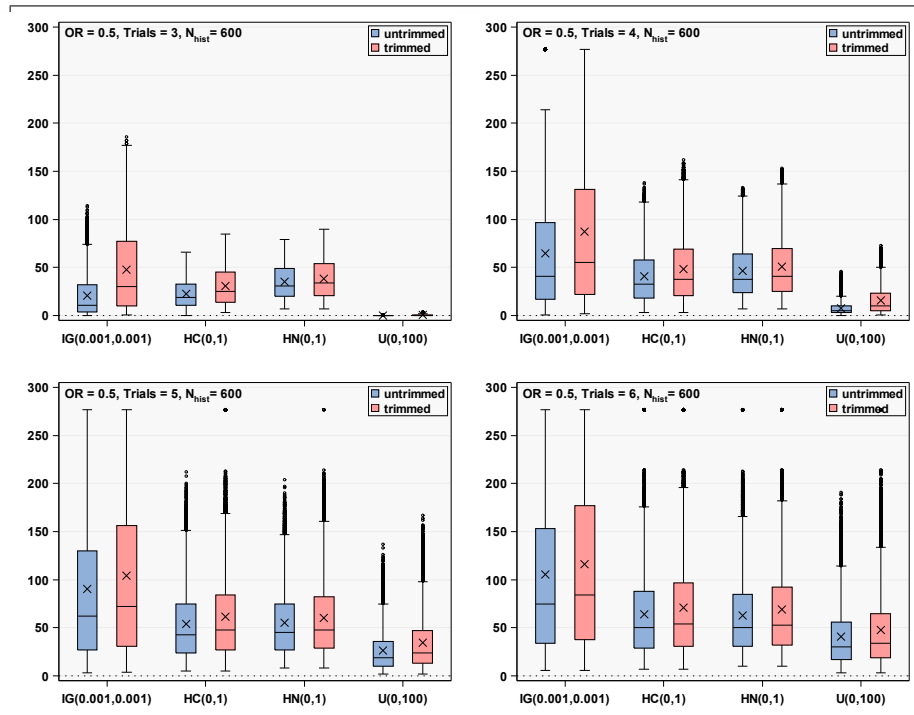
Trials	Prior distribution for τ		Standard								Skewness	
			Mean	Deviation	Minimum	10% Quantil	25% Quantil	Median	75% Quantil	90% Quantil		Maximum
3	$\Gamma^1(0.001,0.001)$	N_{prior}	21.7	35.8	-112	0	2	8	29	60	277	3.5
		N_{prior_trim}	51.7	73.5	-174	1	6	22	71	142	277	1.9
	HC(0,1)	N_{prior}	23.2	26.1	-73	1	6	16	34	57	277	2.3
		N_{prior_trim}	32.4	39.6	-92	1	9	21	46	77	277	2.8
	HN(0,1)	N_{prior}	37.4	41.2	-93	2	12	28	53	83	277	2.5
		N_{prior_trim}	41.9	48.0	-103	2	13	30	58	93	277	2.5
	U(0,100)	N_{prior}	0.2	0.4	-2	0	0	0	0	1	3	1.5
		N_{prior_trim}	0.7	1.1	-5	0	0	0	1	2	8	1.5
4	$\Gamma^1(0.001,0.001)$	N_{prior}	65.3	81.8	-248	3	12	32	88	189	277	1.5
		N_{prior_trim}	82.0	93.9	-307	4	15	43	119	277	277	1.1
	HC(0,1)	N_{prior}	43.7	51.9	-147	3	13	29	58	102	277	2.4
		N_{prior_trim}	52.2	62.0	-179	4	14	33	69	123	277	2.1
	HN(0,1)	N_{prior}	49.7	55.0	-158	5	17	35	66	111	277	2.2
		N_{prior_trim}	55.1	61.7	-173	5	18	38	72	124	277	2.0
	U(0,100)	N_{prior}	7.6	9.7	-52	0	2	4	10	19	97	2.3
		N_{prior_trim}	16.0	21.0	-82	1	4	9	21	42	277	2.9
5	$\Gamma^1(0.001,0.001)$	N_{prior}	86.9	92.1	-262	7	19	51	125	277	277	1.1
		N_{prior_trim}	97.8	98.1	-287	8	22	59	149	277	277	0.9
	HC(0,1)	N_{prior}	58.4	65.1	-166	6	18	38	76	133	277	2.0
		N_{prior_trim}	65.4	71.3	-184	7	19	42	85	152	277	1.7
	HN(0,1)	N_{prior}	59.8	64.0	-158	8	20	40	78	132	277	2.0
		N_{prior_trim}	65.0	68.8	-172	8	21	43	84	147	277	1.8
	U(0,100)	N_{prior}	27.4	35.5	-118	3	7	16	35	65	277	3.4
		N_{prior_trim}	36.6	47.2	-142	3	9	22	47	87	277	2.9
6	$\Gamma^1(0.001,0.001)$	N_{prior}	99.3	96.2	-185	10	25	63	149	277	277	0.9
		N_{prior_trim}	108.0	100.3	-208	11	28	71	172	277	277	0.7
	HC(0,1)	N_{prior}	68.4	71.9	-129	9	22	45	89	157	277	1.7
		N_{prior_trim}	74.7	76.9	-139	10	24	49	99	179	277	1.5
	HN(0,1)	N_{prior}	67.5	69.4	-115	10	23	46	88	149	277	1.8
		N_{prior_trim}	72.8	74.0	-123	10	25	49	95	166	277	1.6
	U(0,100)	N_{prior}	43.8	53.0	-89	5	13	27	56	101	277	2.6
		N_{prior_trim}	51.0	60.8	-106	6	14	31	64	120	277	2.3

Tab. B.14: Deskriptive Statistiken für die Prior Effective Sample Size N_{eff} : Szenario A, $OR = 0.5$.

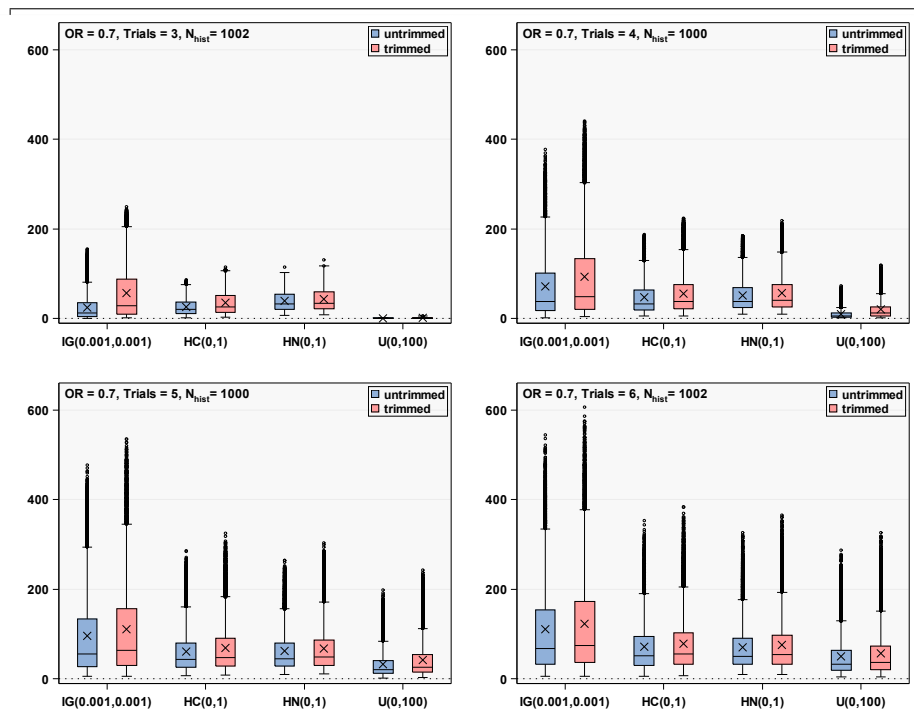
Trials		Prior distribution for τ		Mean	Standard Deviation	Minimum	10% Quantil	25% Quantil	Median	75% Quantil	90% Quantil	Maximum	Skewness
3	$\Gamma^1(0.001, 0.001)$	N_{prior}		27.4	75.9	-238	-8	1	7	31	91	1083	6.6
		N_{prior_trim}		70.1	180.6	-674	-20	1	18	77	215	1083	3.8
	HC(0,1)	N_{prior}		27.0	49.7	-234	-14	2	17	43	85	1083	3.1
		N_{prior_trim}		37.7	76.5	-323	-18	2	22	57	116	1083	4.8
	HN(0,1)	N_{prior}		41.7	75.6	-312	-22	3	30	66	122	1083	4.1
		N_{prior_trim}		47.5	93.7	-356	-24	3	32	73	136	1083	4.9
	U(0,100)	N_{prior}		0.5	0.9	-4	0	0	0	1	2	8	1.0
		N_{prior_trim}		1.2	2.3	-15	0	0	1	2	4	22	1.7
4	$\Gamma^1(0.001, 0.001)$	N_{prior}		82.6	193.1	-666	-19	4	27	91	242	1083	3.5
		N_{prior_trim}		108.2	236.4	-969	-25	5	35	120	323	1083	2.7
	HC(0,1)	N_{prior}		50.4	106.1	-459	-17	4	27	69	143	1083	5.1
		N_{prior_trim}		61.1	131.7	-553	-20	5	31	80	170	1083	4.6
	HN(0,1)	N_{prior}		55.8	110.2	-469	-21	5	34	79	154	1083	4.7
		N_{prior_trim}		62.6	127.1	-540	-22	6	36	86	171	1083	4.5
	U(0,100)	N_{prior}		9.8	20.3	-108	-3	1	5	13	30	231	2.9
		N_{prior_trim}		20.7	46.0	-276	-6	1	9	26	64	1083	5.7
5	$\Gamma^1(0.001, 0.001)$	N_{prior}		109.5	222.1	-896	-19	9	42	124	305	1083	2.9
		N_{prior_trim}		128.1	251.9	-985	-21	11	48	144	358	1083	2.6
	HC(0,1)	N_{prior}		66.0	127.8	-581	-16	9	37	88	178	1083	4.4
		N_{prior_trim}		76.2	151.5	-642	-17	10	40	97	201	1083	4.1
	HN(0,1)	N_{prior}		66.7	124.5	-538	-17	10	40	90	177	1083	4.4
		N_{prior_trim}		74.5	143.8	-610	-18	10	42	97	195	1083	4.2
	U(0,100)	N_{prior}		33.2	66.2	-357	-7	4	17	43	94	1083	5.4
		N_{prior_trim}		44.5	94.5	-501	-10	5	22	55	124	1083	5.5
6	$\Gamma^1(0.001, 0.001)$	N_{prior}		129.2	243.5	-573	-19	13	53	150	350	1083	2.7
		N_{prior_trim}		144.0	264.3	-668	-21	14	58	168	398	1083	2.4
	HC(0,1)	N_{prior}		80.6	150.9	-381	-16	12	44	103	209	1083	4.0
		N_{prior_trim}		89.7	169.1	-417	-17	13	47	112	229	1083	3.7
	HN(0,1)	N_{prior}		78.2	142.6	-362	-16	13	45	101	202	1083	4.1
		N_{prior_trim}		86.2	160.4	-396	-17	13	48	109	220	1083	3.9
	U(0,100)	N_{prior}		54.3	110.1	-314	-10	8	28	67	143	1083	5.3
		N_{prior_trim}		63.2	127.0	-352	-11	8	32	77	167	1083	4.7

Tab. B.15: Deskriptive Statistiken für die Prior Effective Sample Size N_{eff} : Szenario B, $OR = 0.7$.

B.1.2. Adjustierte Simulationsergebnisse



(a) Szenario A: $OR = 0.5$, $\mu_0 = -0.693$.



(b) Szenario B: $OR = 0.7$, $\mu_0 = -0.357$.

Abb. B.12: Boxplots der adjustierten Prior Effective Sample Size N_{prior} .

Für die Berechnung von N_{prior} wurde die a-posteriori prädiktive Varianz des jeweiligen hierarchischen Modells als informativer a-priori Parameter für das Normal-Normal-Modell der neuen Studie benutzt. Der informative a-priori Erwartungswert ist der angenommene Studieneffekt.

Trials	Prior distribution for τ		Standard	10%	25%	75%	90%	Maximum	Skewness			
			Mean	Deviation	Quantil	Quantil	Quantil			Quantil		
3	$\Gamma^{-1}(0.001, 0.001)$	N_{prior} with $\mu_0 = -0.693$	20.6	21.7	0	2	4	11	32	55	114	1.3
		N_{prior_trim} with $\mu_0 = -0.693$	47.7	45.1	1	5	10	30	77	122	186	1.0
	HC(0,1)	N_{prior} with $\mu_0 = -0.693$	22.6	14.4	0	7	11	19	33	45	66	0.7
		N_{prior_trim} with $\mu_0 = -0.693$	30.5	19.7	3	9	14	25	45	62	85	0.7
	HN(0,1)	N_{prior} with $\mu_0 = -0.693$	35.0	18.0	7	14	20	31	49	63	79	0.5
		N_{prior_trim} with $\mu_0 = -0.693$	38.6	20.7	7	15	21	34	54	71	90	0.6
	U(0,100)	N_{prior} with $\mu_0 = -0.693$	0.1	0.3	0	0	0	0	0	1	1	2.5
		N_{prior_trim} with $\mu_0 = -0.693$	0.6	0.8	0	0	0	0	1	2	4	1.0
4	$\Gamma^{-1}(0.001, 0.001)$	N_{prior} with $\mu_0 = -0.693$	64.7	62.2	1	9	17	41	97	156	277	1.4
		N_{prior_trim} with $\mu_0 = -0.693$	87.4	82.9	2	12	22	55	131	277	277	1.1
	HC(0,1)	N_{prior} with $\mu_0 = -0.693$	41.1	28.3	3	12	18	33	58	85	138	0.9
		N_{prior_trim} with $\mu_0 = -0.693$	48.5	34.3	3	14	21	38	69	103	162	1.0
	HN(0,1)	N_{prior} with $\mu_0 = -0.693$	46.3	28.1	7	17	24	38	64	90	133	0.9
		N_{prior_trim} with $\mu_0 = -0.693$	50.9	32.1	7	18	25	41	70	101	153	0.9
	U(0,100)	N_{prior} with $\mu_0 = -0.693$	7.6	7.0	0	2	3	5	10	18	46	1.7
		N_{prior_trim} with $\mu_0 = -0.693$	15.9	14.4	1	3	5	10	23	38	73	1.4
5	$\Gamma^{-1}(0.001, 0.001)$	N_{prior} with $\mu_0 = -0.693$	90.3	80.2	3	15	27	62	130	212	277	1.1
		N_{prior_trim} with $\mu_0 = -0.693$	104.4	89.0	4	17	31	72	156	277	277	0.9
	HC(0,1)	N_{prior} with $\mu_0 = -0.693$	54.2	38.3	5	16	24	43	75	111	212	1.1
		N_{prior_trim} with $\mu_0 = -0.693$	61.3	44.9	5	17	27	48	84	127	277	1.3
	HN(0,1)	N_{prior} with $\mu_0 = -0.693$	55.4	36.0	8	19	27	45	75	109	204	1.1
		N_{prior_trim} with $\mu_0 = -0.693$	60.5	40.8	8	20	29	48	82	121	277	1.2
	U(0,100)	N_{prior} with $\mu_0 = -0.693$	26.3	22.1	2	6	10	19	36	59	137	1.5
		N_{prior_trim} with $\mu_0 = -0.693$	34.5	29.2	2	8	13	24	47	77	167	1.5
6	$\Gamma^{-1}(0.001, 0.001)$	N_{prior} with $\mu_0 = -0.693$	105.2	87.2	6	19	34	75	153	277	277	0.9
		N_{prior_trim} with $\mu_0 = -0.693$	116.1	92.4	6	21	38	84	177	277	277	0.7
	HC(0,1)	N_{prior} with $\mu_0 = -0.693$	64.2	46.4	7	19	29	50	88	132	277	1.4
		N_{prior_trim} with $\mu_0 = -0.693$	71.2	53.7	7	20	31	54	97	147	277	1.5
	HN(0,1)	N_{prior} with $\mu_0 = -0.693$	63.1	42.8	10	21	31	50	85	126	277	1.3
		N_{prior_trim} with $\mu_0 = -0.693$	68.8	49.4	10	22	32	53	92	139	277	1.5
	U(0,100)	N_{prior} with $\mu_0 = -0.693$	40.9	32.7	3	11	17	30	56	89	191	1.4
		N_{prior_trim} with $\mu_0 = -0.693$	47.6	38.9	3	12	19	34	65	105	277	1.5

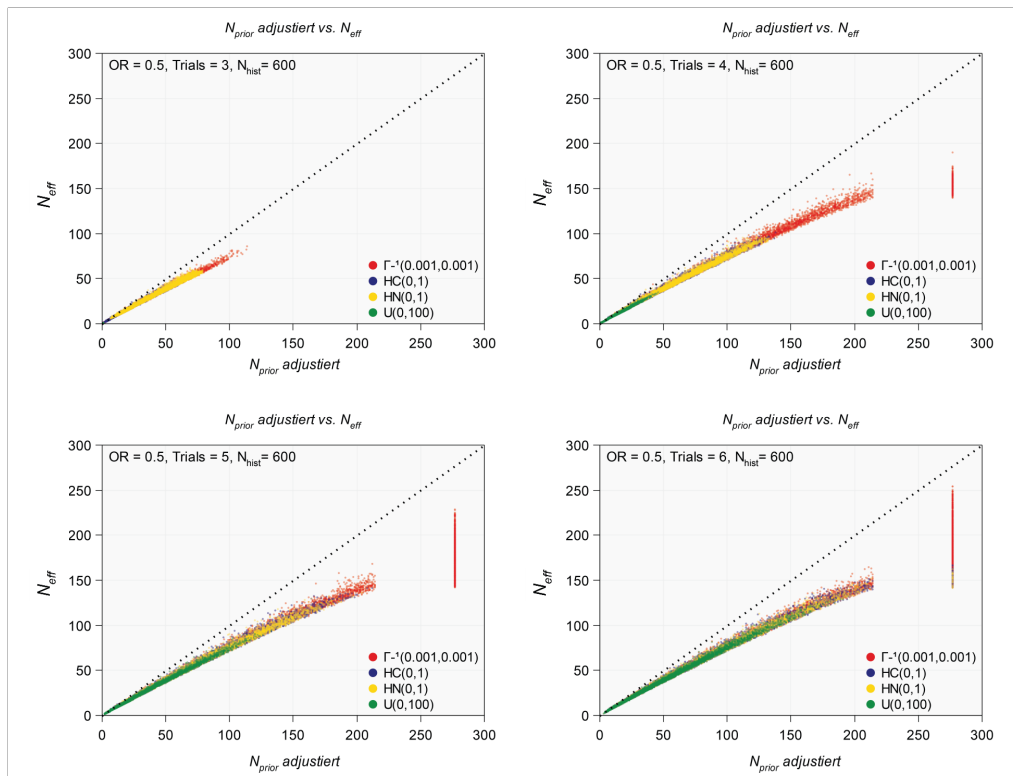
Tab. B.16: Deskriptive Statistiken für die adjustierte Prior Effective Sample Size N_{prior} : Szenario A, $OR = 0.5$.

Für die Berechnung von N_{prior} wurde die a-posteriori prädiktive Varianz des jeweiligen hierarchischen Modells als informativer a-priori Parameter für das Normal-Normal-Modell der neuen Studie benutzt. Der informative a-priori Erwartungswert ist der angenommene Studieneffekt $\mu_0 = -0.693$.

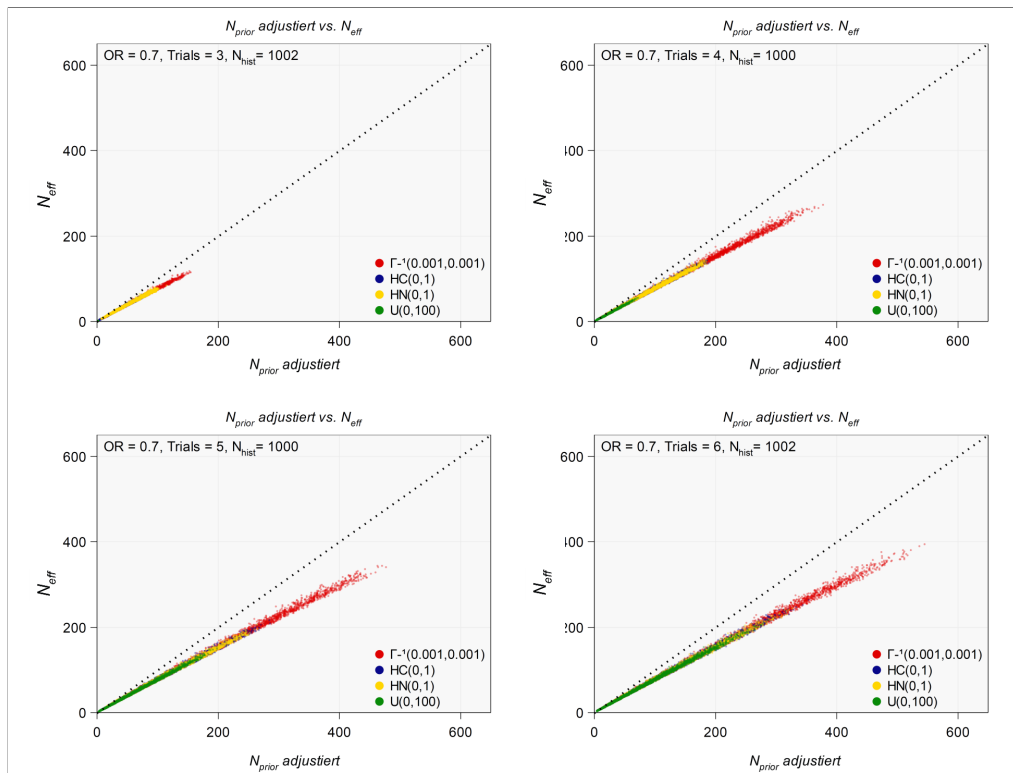
Trials	Prior distribution for τ		Standard		10%	25%	75%	90%	Maximum	Skewness		
			Mean	Deviation	Minimum	Quantil	Quantil	Median			Quantil	
3	$\Gamma^{-1}(0.001, 0.001)$	N_{prior} with $\mu_0 = -0.357$	25.0	29.8	0	2	4	12	35	71	155	1.7
		N_{prior_trim} with $\mu_0 = -0.357$	56.8	61.3	1	5	10	28	88	163	250	1.2
	HC(0,1)	N_{prior} with $\mu_0 = -0.357$	26.0	18.5	1	8	11	20	37	56	86	1.0
		N_{prior_trim} with $\mu_0 = -0.357$	35.2	25.8	3	10	14	26	51	78	115	1.0
	HN(0,1)	N_{prior} with $\mu_0 = -0.357$	39.1	22.9	7	15	20	32	54	77	115	0.8
		N_{prior_trim} with $\mu_0 = -0.357$	43.3	26.7	8	16	22	34	60	87	131	0.9
	U(0,100)	N_{prior} with $\mu_0 = -0.357$	0.4	0.5	0	0	0	0	1	1	2	0.4
		N_{prior_trim} with $\mu_0 = -0.357$	1.2	1.1	0	0	0	1	2	3	6	1.1
4	$\Gamma^{-1}(0.001, 0.001)$	N_{prior} with $\mu_0 = -0.357$	71.2	75.4	2	10	17	38	101	192	378	1.5
		N_{prior_trim} with $\mu_0 = -0.357$	93.1	97.9	4	13	21	49	134	255	441	1.4
	HC(0,1)	N_{prior} with $\mu_0 = -0.357$	46.9	37.4	5	13	19	33	63	105	188	1.4
		N_{prior_trim} with $\mu_0 = -0.357$	55.2	45.5	5	15	22	38	75	127	224	1.4
	HN(0,1)	N_{prior} with $\mu_0 = -0.357$	51.8	36.9	9	18	24	38	69	110	185	1.3
		N_{prior_trim} with $\mu_0 = -0.357$	57.1	42.5	9	19	26	41	75	123	219	1.4
	U(0,100)	N_{prior} with $\mu_0 = -0.357$	9.9	10.3	1	2	3	6	12	24	73	2.2
		N_{prior_trim} with $\mu_0 = -0.357$	20.4	21.6	1	4	6	12	26	53	119	1.9
5	$\Gamma^{-1}(0.001, 0.001)$	N_{prior} with $\mu_0 = -0.357$	95.7	95.4	6	16	27	55	134	247	477	1.5
		N_{prior_trim} with $\mu_0 = -0.357$	111.3	111.7	6	18	30	63	156	289	536	1.5
	HC(0,1)	N_{prior} with $\mu_0 = -0.357$	61.4	50.0	7	18	26	43	80	136	286	1.6
		N_{prior_trim} with $\mu_0 = -0.357$	68.8	57.4	8	19	28	47	90	155	325	1.6
	HN(0,1)	N_{prior} with $\mu_0 = -0.357$	62.2	47.3	10	21	29	45	80	133	265	1.6
		N_{prior_trim} with $\mu_0 = -0.357$	67.8	53.5	11	22	30	48	87	147	303	1.6
	U(0,100)	N_{prior} with $\mu_0 = -0.357$	32.6	31.6	2	8	12	20	41	77	198	1.9
		N_{prior_trim} with $\mu_0 = -0.357$	42.3	41.5	3	10	15	26	54	101	243	1.9
6	$\Gamma^{-1}(0.001, 0.001)$	N_{prior} with $\mu_0 = -0.357$	110.0	105.3	5	21	33	67	154	275	545	1.5
		N_{prior_trim} with $\mu_0 = -0.357$	123.2	119.0	6	23	36	74	173	311	607	1.4
	HC(0,1)	N_{prior} with $\mu_0 = -0.357$	71.7	58.2	6	21	30	51	94	160	354	1.6
		N_{prior_trim} with $\mu_0 = -0.357$	78.5	65.0	7	23	33	55	102	177	384	1.6
	HN(0,1)	N_{prior} with $\mu_0 = -0.357$	70.3	54.6	9	23	32	50	90	152	327	1.6
		N_{prior_trim} with $\mu_0 = -0.357$	76.0	60.7	9	24	33	54	97	166	365	1.6
	U(0,100)	N_{prior} with $\mu_0 = -0.357$	49.5	45.3	4	13	19	32	63	115	287	1.8
		N_{prior_trim} with $\mu_0 = -0.357$	57.1	53.0	4	14	21	37	73	133	326	1.8

Tab. B.17: Deskriptive Statistiken für die adjustierte Prior Effective Sample Size N_{prior} : Szenario B, $OR = 0.7$.

Für die Berechnung von N_{prior} wurde die a-posteriori prädiktive Varianz des jeweiligen hierarchischen Modells als informativer a-priori Parameter für das Normal-Normal-Modell der neuen Studie benutzt. Der informative a-priori Erwartungswert ist der angenommene Studieneffekt $\mu_0 = -0.357$.



(a) Szenario A: $OR = 0.5$.



(b) Szenario B: $OR = 0.7$.

Abb. B.13: Scatterplots: Zusammenhang zwischen N_{prior} adjustiert und N_{eff} .

Für die Berechnung von N_{prior} wurde die a-posteriori prädiktive Varianz des vollen hierarchischen Modells als informativer a-priori Parameter für das Normal-Normal-Modell der neuen Studie benutzt. Der informative a-priori Erwartungswert ist der angenommene Studieneffekt $\mu_0 = -0.693$ bzw. $\mu_0 = -0.357$.

C Programmsyntax

C.1 WinBUGS/OpenBUGS Code der Anwendungsbeispiele

C.1.1. Code für die Heparin-Studien

Der WinBUGS bzw. OpenBUGS Code für die Heparin-Studien ist exemplarisch für die Inverse-Gamma-Verteilung $\Gamma^{-1}(0.001,0.001)$ dargestellt. Die Konstruktion der anderen a-priori Verteilungen kann in der R-Syntax für die Simulationsstudie C.3 nachgeschlagen werden.

```
# Daten
list(
  A=1, B=18, C=19, D=19,
  n.trt = c( 98,202, 75,85,213,304, 42,195,29,247,610, 68,101,388,510,265,125,627,36),
  ev.trt = c( 6, 14,0.5, 6, 6, 5,0.5, 10, 1, 13, 21,0.5, 5, 7, 27, 6, 2, 22, 2),
  n.ctrl = c(196,198, 73,85,219,308, 40,205,30,253,290, 68,103,375,511,273,124,593,40),
  ev.ctrl =c( 13, 17,3.5,12, 15, 6,1.5, 12, 3, 17, 11,3.5, 3, 24, 25, 15, 4, 27, 4)
)

# Startwerte für Markovkette 1
list(mu=0, shift=0, Prec.theta=0.2)

# Startwerte für Markovkette 2
list(mu=-0.2, shift=-0.2, Prec.theta=0.5)

# OpenBUGS/WinBUGS Modell
model{
  # Datenmodell der Erwachsenen
  for (a in A:B)
  {
    # Sampling Model (Likelihood)
    Y[a] <- log(ev.trt[a] * (n.ctrl[a]-ev.ctrl[a]) / (ev.ctrl[a] * (n.trt[a]-ev.trt[a])))
    VARlnOR[a] <- (1/ev.trt[a]) + (1/(n.trt[a]-ev.trt[a])) + (1/ev.ctrl[a]) + (1/(n.ctrl[a]-ev.ctrl[a]))
    PREClNOR[a] <- 1/VARlnOR[a]
    Y[a] ~ dnorm(theta[a],PREClNOR[a])

    # Parameter Model
    theta[a] ~ dnorm(mu,Prec.theta)
  }

  # Datenmodell der Kinder
  for (b in C:D)
  {
    # Sampling Model (Likelihood)
    Y[b] <- log( ev.trt[b]*(n.ctrl[b]-ev.ctrl[b])/(ev.ctrl[b]*(n.trt[b]-ev.trt[b])))
    VARlnOR[b] <- (1/ev.trt[b]) + (1/(n.trt[b]-ev.trt[b])) + (1/ev.ctrl[b]) + (1/(n.ctrl[b]-ev.ctrl[b]))
    PREClNOR[b] <- 1/VARlnOR[b]
    Y[b] ~ dnorm(theta[b],PREClNOR[b])

    # Parameter Model
    theta[b] ~ dnorm(mushift,Prec.theta)
  }

  mushift <- mu + shift

  #####
  # A-priori Verteilungen
  mu ~ dnorm(0,1.0E-6)
  shift ~ dnorm(0,1.0E-6)

  # A-priori Verteilung für tau^2 InvGamma (0.001,0.001)
  Prec.theta ~ dgamma(0.001,0.001)
  tau <- 1/(sqrt(Prec.theta))

  #####
  # Vorhersage Parameter für eine neue Studie
  theta.new.adult ~ dnorm(mu,Prec.theta)
  theta.new.paed ~ dnorm(mushift,Prec.theta)
}
```


C.1.2. Code für die Metformin-Studien

Der WinBUGS bzw. OpenBUGS Code für die Metformin-Studien ist exemplarisch für die Gleichverteilung $U(0,100)$ dargestellt. Die Konstruktion der anderen a-priori Verteilungen kann in der R-Syntax für die Simulationsstudie C.3 nachgeschlagen werden.

```
# Daten
list(
  A=1, B=28, C=29, D=29,
  MW =c( -0.74, -1.77,  0.05, -0.80, -3.98, -1.87, -0.30, -1.06, -0.47, -0.63, -2.42,
        -0.30, -0.80, -2.29, -0.62, -1.39, -0.69, -0.32, -0.84, -0.29, -0.30, -0.33,
        -1.66, -1.85, -0.75, -0.65, -0.54, -0.42, -0.83),
  Std =c( 0.276, 0.347, 0.163, 0.163, 0.750, 0.592, 0.097, 0.128, 0.087, 0.265, 0.347,
        0.163, 0.204, 0.520, 0.240, 0.286, 0.112, 0.367, 0.306, 0.265, 0.276, 0.260,
        0.163, 0.321, 0.347, 0.204, 0.439, 0.077, 0.087)
)

# Startwerte für Markovkette 1
list(mu=0, shift=0,tau=1)

# Startwerte für Markovkette 2
list(mu=-0.2, shift=-0.2,tau=0.5)

# OpenBUGS/WinBUGS Modell
model{
  # Datenmodell der Erwachsenen
  for (a in A:B)
  {
    # Sampling Model (Likelihood)
    Y[a] <- MW[a]
    VARMw[a] <- Std[a]*Std[a]
    PRECMw[a] <- 1/VARMw[a]
    Y[a] ~ dnorm(theta[a],PRECMw[a])

    # Parameter Model
    theta[a] ~ dnorm(mu, Prec.theta)
  }

  # Datenmodell der Kinder
  for (b in C:D)
  {
    # Sampling Model (Likelihood)
    Y[b] <- MW[b]
    VARMw[b] <- Std[b]*Std[b]
    PRECMw[b] <- 1/VARMw[b]
    Y[b] ~ dnorm(theta[b],PRECMw[b])

    # Parameter Model
    theta[b] ~ dnorm(mushift, Prec.theta)
  }

  mushift <- mu + shift

  #####
  # A-priori Verteilungen
  mu ~ dnorm(0,1.0E-6)
  shift ~ dnorm(0,1.0E-6)

  # A-priori Verteilung für tau Gleichverteilung U(1.0E-6,100)
  tau ~ dunif(1.0E-6, 100)
  Prec.theta <- 1/(tau*tau)

  #####
  # Vorhersage Parameter für eine neue Studie
  theta.new.adult ~ dnorm(mu, Prec.theta)
  theta.new.paed ~ dnorm(mushift, Prec.theta)
}
```

C.1.3. Code für die FeNO-studien

Der WinBUGS bzw. OpenBUGS Code für die Metformin-Studien ist exemplarisch für die Half-Normal-Verteilung $HN(0,1)$ dargestellt. Die Konstruktion der anderen a-priori Verteilungen kann in der R-Syntax für die Simulationsstudie C.3 nachgeschlagen werden.

```
# Daten
# data=list(
```

```

A=1, B=2, C=3, D=5,
n.trt =c(52, 46, 75, 42, 276),
ev.trt =c(12, 14, 9, 7, 102),
n.ctrl =c(51, 48, 72, 47, 270),
ev.ctrl=c(19, 11, 12, 10, 118 )
)

# Startwerte für Markovkette 1
list(mu=0, shift=0,tau=1)

# Startwerte für Markovkette 2
inits2=list(mu=-0.2, shift=-0.2,tau=0.5)

# OpenBUGS/WinBUGS Modell
modell{
# Datenmodell der Erwachsenen
for (a in A:B)
{
# Sampling Model (Likelihood)
Y[a] <- log(ev.trt[a] * (n.ctrl[a]-ev.ctrl[a])) / (ev.ctrl[a] * (n.trt[a]-ev.trt[a]))
VARlnOR[a] <- (1/ev.trt[a]) + (1/(n.trt[a]-ev.trt[a])) + (1/ev.ctrl[a]) + (1/(n.ctrl[a]-ev.ctrl[a]))
PREClNOR[a] <- 1/VARlnOR[a]
Y[a] ~ dnorm(theta[a],PREClNOR[a])

# Parameter Model
theta[a] ~ dnorm(mu,PREC.theta)
}

# Datenmodell der Kinder
for (b in C:D)
{
# Sampling Model (Likelihood)
Y[b] <- log( ev.trt[b]*(n.ctrl[b]-ev.ctrl[b])/(ev.ctrl[b]*(n.trt[b]-ev.trt[b])))
VARlnOR[b] <- (1/ev.trt[b]) + (1/(n.trt[b]-ev.trt[b])) + (1/ev.ctrl[b]) + (1/(n.ctrl[b]-ev.ctrl[b]))
PREClNOR[b] <- 1/VARlnOR[b]
Y[b] ~ dnorm(theta[b],PREClNOR[b])

# Parameter Model
theta[b] ~ dnorm(mushift,PREC.theta)
}

mushift <- mu + shift

#####
# A-priori Verteilungen
mu ~ dnorm(0,1.0E-6)
shift ~ dnorm(0,1.0E-6)

# A-priori Verteilung für tau Half-Normal-Verteilung H\!C(0,1)
tau ~ dnorm(0,1)I(0,)
PREC.theta <- 1/(tau*tau)

#####
# Vorhersage Parameter für eine neue Studie
theta.new.adult ~ dnorm(mu,PREC.theta)
theta.new.paed ~ dnorm(mushift,PREC.theta)
}

```

C.2 R-Skript zur Berechnung von N_{prior} für die Heparin- und FeNO-Studien

Die Ergebnisse der Berechnungen sind in den Tabellen 4.3 und 4.7 dargestellt.

```
#####
# Funktion zur Bestimmung von Nref im Normal-Normal-Modell mit nicht-informativen a-priori Parametern
#####
Power.uninformativ <- function(Nref)
{
  #Nref=0
  power.uninf <-0
  while(power.uninf < 0.8){
    Nref <- Nref + 1
    npergroup <- Nref/2

    ln.or <- log( (pT/(1-pT)) / (pC/(1-pC)) )
    sigma <- sqrt( 1/(pT*(1-pT)*npergroup) + 1/(pC*(1-pC)*npergroup) )
    power.uninf <- pnorm(qnorm(0.025)-(ln.or /sigma)
    )
  }

  rm(ln.or, sigma)
  return(list(power.uninf, Nref))
}

#####
# Funktion zur Bestimmung von Nprior im Normal-Normal-Modell unter Benutzung informativer a-priori Parameter
#####
Nprior.informativ <- function(Nref)
{
  Nref.inf <- 0
  power <- 0 # Variable muss für while-Schleife vorher existieren
  while(power < 0.8){

    # Nref.inf hier dann für jeden Durchlauf um 1 hochsetzen
    Nref.inf <- Nref.inf +1
    npergroup <- Nref.inf/2

    ln.or <- log( (pT/(1-pT)) / (pC/(1-pC)) )
    sigma <- sqrt( 1/(pT*(1-pT)*npergroup) + 1/(pC*(1-pC)*npergroup) )
    power <- pnorm( ( (sqrt(Sd.logORnew^(-2)+sigma^(-2))*qnorm(0.025) - Sd.logORnew^(-2)*(E.logOR))*sigma^2 - ln.or )
    / sigma )
  }
  Nprior <- Nref - Nref.inf
  ratioNpriorNref <- round(Nprior/Nref, digits = 3)
  ratioNpriorNhist <- round(Nprior/Nhist, digits = 3)
  return(list(power, Nprior, ratioNpriorNref, ratioNpriorNhist))
  rm(E.logOR, Sd.logORnew, ln.or, sigma, power)
}

#####
#####
#####
# Heparin-Studien
#####
Nhist <- 8198 # historische Gesamtfallzahl
#####
# KINDERSTUDIEN

pT <- 2/36 # Ergebnisse der Kinderstudie
pC <- 4/40 # Ergebnisse der Kinderstudie

# Fallzahlplanung nicht-informativ
Power.uninformativ(0)
Nref <- Power.uninformativ(0)[[2]] # Nref=1171
Nref <- 1172
#Nref <- 1130

#####
# Bestimmung von Nprior informativ
#####

# Inv-Gamma(0.001,0.001)
# untrimmed
E.logOR <- -0.663
Sd.logORnew <- 0.937
Nprior.informativ(Nref)

# trimmed
E.logOR <- -0.663
Sd.logORnew <- 0.926
Nprior.informativ(Nref)
#####
# HN(0,1)
# untrimmed
E.logOR <- -0.615
Sd.logORnew <- 0.966
Nprior.informativ(Nref)

# trimmed
```

```

E.logOR <- -0.615
Sd.logORnew <- 0.954
Nprior.informativ(Nref)
#####
# U(0,100)
# untrimmed
E.logOR <- -0.634
Sd.logORnew <- 0.989
Nprior.informativ(Nref)

# trimmed
E.logOR <- -0.639
Sd.logORnew <- 0.978
Nprior.informativ(Nref)
#####
# HC(0,1)
# untrimmed
E.logOR <- -0.667
Sd.logORnew <- 0.957
Nprior.informativ(Nref)

# trimmed
E.logOR <- -0.667
Sd.logORnew <- 0.946
Nprior.informativ(Nref)

#####
#####
#####
# ERWACHSENENSTUDIEN

# pT <- 151/4181 # Ergebnisse der Erwachsenenstudien
# pC <- 211/3941 # Ergebnisse der Erwachsenenstudien
pTC <- (151+211)/(4181+3941)
pT <- 0.0364396 # Berechnet zu logOR=-0.386 bei angenommener gemeinsamer Ereigniswahrscheinlichkeit pTC
pC <- 0.052701

# Fallzahlplanung nicht-informativ, Heparin-Studie an Erwachsenen
Power.uninformativ(0)
Nref <- Power.uninformativ(0)[[2]]
# Nref <- 5112

#####
# Bestimmung von Nprior informativ
#####

# Inv-Gamma(0.001,0.001)
# untrimmed
E.logOR <- -0.375
Sd.logORnew <- 0.2411
Nprior.informativ(Nref)

# trimmed
E.logOR <- -0.375
Sd.logORnew <- 0.2407
Nprior.informativ(Nref)
#####
# HN(0,1)
# untrimmed
E.logOR <- -0.386
Sd.logORnew <- 0.3070
Nprior.informativ(Nref)

# trimmed
E.logOR <- -0.386
Sd.logORnew <- 0.3063
Nprior.informativ(Nref)
#####
# U(0,100)
# untrimmed
E.logOR <- -0.386
Sd.logORnew <- 0.3118
Nprior.informativ(Nref)

# trimmed
E.logOR <- -0.386
Sd.logORnew <- 0.3111
Nprior.informativ(Nref)
#####
# HC(0,1)
# untrimmed
E.logOR <- -0.384
Sd.logORnew <- 0.2937
Nprior.informativ(Nref)

# trimmed
E.logOR <- -0.384
Sd.logORnew <- 0.2931
Nprior.informativ(Nref)

#####
#####

```

```
#####
# FeNO-Studien
#####
Nhist <- 979 # historische Gesamtfallzahl
#####
# KINDERSTUDIEN

pC <- (12+10+118)/(72+47+270) # FeNO-Ergebnisse der Kinderstudien
pT <- (9+7+102)/(75+42+276) # FeNO-Ergebnisse der Kinderstudien

# Fallzahlplanung nicht-informativ, FeNO-Studie an Kindern
Power.uninformativ(0)
Nref <- Power.uninformativ(0)[[2]]
# Nref <- 1956

#####
# Bestimmung von Nprior informativ
#####

# Inv-Gamma(0.001,0.001)
# untrimmed
E.logOR <- -0.301
Sd.logORnew <- 0.541
Nprior.informativ(Nref)

# trimmed
E.logOR <- -0.302
Sd.logORnew <- 0.455
Nprior.informativ(Nref)
#####
# HN(0,1)
# untrimmed
E.logOR <- -0.308
Sd.logORnew <- 0.637
Nprior.informativ(Nref)

# trimmed
E.logOR <- -0.308
Sd.logORnew <- 0.608
Nprior.informativ(Nref)
#####
# U(0,100)
# untrimmed
E.logOR <- -0.310
Sd.logORnew <- 1.785
Nprior.informativ(Nref)

# trimmed
E.logOR <- -0.308
Sd.logORnew <- 1.146
Nprior.informativ(Nref)
#####
# HC(0,1)
# untrimmed
E.logOR <- -0.308
Sd.logORnew <- 0.676
Nprior.informativ(Nref)

# trimmed
E.logOR <- -0.308
Sd.logORnew <- 0.625
Nprior.informativ(Nref)

#####
#####
#####
#####
# ERWACHSENENSTUDIEN

# pT <- 26/98 # FeNO-Ergebnisse der Erwachsenenstudien
# pC <- 30/89 # FeNO-Ergebnisse der Erwachsenenstudien

pTC <- (26+30)/(98+89)
pT <- 0.28144 # Berechnet für lorOR=0.172 bei angenommener mittlerer Ereigniswahrscheinlichkeit pTC
pC <- 0.3174904

# Fallzahlplanung nicht-informativ, FeNO-Studie an Erwachsenenstudien
Power.uninformativ(0)
Nref <- Power.uninformativ(0)[[2]]
# Nref <- 5073
Nref <- 5074

#####
# Bestimmung von Nprior informativ
#####

# Inv-Gamma(0.001,0.001)
# untrimmed
E.logOR <- -0.177
Sd.logORnew <- 0.599
Nprior.informativ(Nref)
```

```
# trimmed
E.logOR <- -0.177
Sd.logORnew <- 0.578
Nprior.informativ(Nref)
#####
# HN(0,1)
# untrimmed
E.logOR <- -0.172
Sd.logORnew <- 0.713
Nprior.informativ(Nref)

# trimmed
E.logOR <- -0.172
Sd.logORnew <- 0.708
Nprior.informativ(Nref)
#####
# U(0,100)
# untrimmed
E.logOR <- -0.171
Sd.logORnew <- 1.913
Nprior.informativ(Nref)

# trimmed
E.logOR <- -0.167
Sd.logORnew <- 1.461
Nprior.informativ(Nref)
#####
# HC(0,1)
# untrimmed
E.logOR <- -0.172
Sd.logORnew <- 0.758
Nprior.informativ(Nref)

# trimmed
E.logOR <- -0.172
Sd.logORnew <- 0.741
Nprior.informativ(Nref)
```

C.3 R-Skript der Simulationsstudie

```

setwd("d:/IBKF/Bayes/Koch/Simulation/Simulierte Studien/") # Arbeitsverzeichnis festlegen
OpenBUGSdirectory <- "c:/Programme/OpenBUGS/OpenBUGS" # Installationsverzeichnis von OpenBUGS festlegen
#WinBUGSdirectory <- "c:/Programme/WinBUGS/" # Installationsverzeichnis von WINBUGS festlegen

outputcsvname <- "Setting1trials3.csv" # Name der *.csv Ergebnisdatei festlegen

library(gtools)
library(R2WinBUGS)

#####
#####
# HILFSFUNKTIONEN #####
#####
# Funktionen zur Genierung von Kontingenztafeln für die jeweiligen einzelnen Studien
#
# Zu vorgegebenem Odds Ratio, mittlerer Ereignisrate und Fallzahl wird eine Kontingenztafel
# erzeugt, deren Ereignisse ein OR erzeugen, dass so nah wie möglich an dem vorgegebenem
# Odds Ratio liegt.

# Funktion zur Berechnung von odds (auch für p=1)
odd <- function(p){
  if(p == 1){temp <- Inf}
  if(p != 1){temp <- p / (1-p)}
  return(temp)
}

# Inverse Funktion der Funktion zur Berechnung von odds (auch für y = unendlich = Inf)
invodd <- function(y){
  if(y == Inf){temp <- 1}
  if(y != Inf){temp <- y/(1+y)}
  return(temp)
}

# Funktion zur Berechnung von OR in Abhängigkeit von eT
OR.Berechnung <- function(temp.eT){
  pT <- temp.eT / nT
  eC <- eTC - temp.eT
  pC <- eC / nC
  return(odd(pT) / odd(pC))
}

# Funktion zur Berechnung der absoluten Differenz von OR in Abhängigkeit von eT
OR.difference <- function(temp.eT){
  pT <- temp.eT / nT
  eC <- eTC - temp.eT
  pC <- eC / nC
  return(abs(odd(pT) / odd(pC) - OR))
}

# Funktion zur Berechnung von eT in Abhängigkeit von nT, nC, eTC und OR
eT.Berechnung <- function(nT,nC,eTC,OR){

  N <- nT + nC
  #eT <- (nT*OR*eTC) / (OR*(nT+1) + nC -1)

  Nullstellenfunktion.eT <- function(temp.eT) {
    return(temp.eT + invodd(odd(temp.eT/nT) / OR)*nC - eTC)
  }

  eT.vorläufig <- uniroot(Nullstellenfunktion.eT, c(0,nT))["root"]

  #ganzzahlige Nachbarn von eT.vorläufig
  eTl <- floor(eT.vorläufig)
  eTu <- ceiling(eT.vorläufig)

  #Wähle eTl bzw. eTu so, dass OR.difference klein ist
  eT <- (OR.difference(eTl) < OR.difference(eTu)) * eTl + (OR.difference(eTl) >= OR.difference(eTu)) * eTu
  return(eT)
}

# Funktion zur Erstellung der Kontingenztafel in Abhängigkeit von nT, nC, eTC und OR
Kontingenztafel <- function(nT,nC,eTC,OR){
  eT <- eT.Berechnung(nT,nC,eTC,OR)
  eC <- eTC - eT
  nicht.eT <- nT - eT
  nicht.eC <- nC - eC

  N <- nT + nC

  odd.T <- eT/nicht.eT
  odd.K <- eC/nicht.eC
  OR.temp <- odd.T / odd.K

  Resultat <- data.frame(EreignisT=eT,KeinEreignisT=nicht.eT,EreignisK=eC,KeinEreignisK=nicht.eC,
    OR.berechnet = OR.temp, OR.gewünscht = OR, AnzahlEreignisse = eTC,
    Treat = nT, Kontrolle = nC, Stichprobenumfang = N)
  return(Resultat)
}

# Berechnung der Anzahl von Ereignissen in der Treat-Gruppe in Abhängigkeit von nT, nC, OR und eTC
#eT.Berechnung(nT,nC,eTC,OR)

# Berechnung des Odds-Ratios in Abhängigkeit von eT (und in Abhängigkeit von nT, nC, OR und eTC)
#OR.Berechnung(eT)

```

```

# Berechnung der absoluten Differenz zwischen gewünschtem Odds-Ratio und dem berechneten
# Odds-Ratio in Abhängigkeit von eT (und in Abhängigkeit von nT, nC, OR und eTC)
#OR.difference(eT)

# Erzeugung Kontingenztafeln in Abhängigkeit von nT, nC, OR und eTC, so dass der Abstand
# zwischen dem gewünschten und berechneten Odds-Ratio minimal ist.
#Kontingenztafel(nT,nC,eTC,OR)

#####
#####
# SIMULATIONSSZENARIEN FESTLEGEN #####
#####

#####
# SETTING 1
# OR=0.5, Standardabweichung zwischen den Studien=0.5
# N_hist=600, Studien pro Metaanalyse 3,4,5,6

# Feste Hyperparameter: Die Wahrheit!
OR_0 <- 0.5 # bekanntes wahre Odds-Ratio
mu_0 <- log(OR_0)
tau_0 <- 0.5 # bekannte wahre Standardabweichung zwischen den einzelnen Studien

# Wahre Ereigniswahrscheinlichkeiten Kontrollgruppe pC und der Behandlungsgruppe pT
pT <- 1/3
pC <- 0.5
pTC <- (pC+pT)/2 # mittlere Ereigniswahrscheinlichkeit

# Gesamtfallzahl der einzelnen Studien
n.trials <- 200
#n.trials <- 150
#n.trials <- 120
#n.trials <- 100

# Simulierte Fallzahl mit nicht-informativer hyper a-priori Verteilung
# für eine Power=80%, mit pT <- 1/3, pC <- 0.5 , OR=0.5
N_ref <- 270

#####
###SETTING 2
#OR=0.7, Standardabweichung zwischen den Studien=0.5
#N_hist=1000/1002, Studien pro Metaanalyse 3,4,5,6

#Feste Hyperparameter: Die Wahrheit!
# OR_0 <- 0.7 # bekanntes wahre Odds-Ratio
# mu_0 <- log(OR_0)
# tau_0 <- 0.5 # bekannte wahre Standardabweichung zwischen den einzelnen Studien

#Wahre Ereigniswahrscheinlichkeiten Kontrollgruppe pC und der Behandlungsgruppe pT
# pT <- 0.3182
# pC <- 0.4
# pTC <- (pC+pT)/2 # mittlere Ereigniswahrscheinlichkeit

##Gesamtfallzahl der einzelnen Studien
#n.trials <- 334
#n.trials <- 250
#n.trials <- 200
# n.trials <- 167

# Simulierte Fallzahl mit nicht-informativer hyper a-priori Verteilung
##für eine Power=80%, mit pT <- 0.3182, pC <- 0.4 , OR=0.7
# N_ref <- 1074

#####
#####
# EINSTELLEN!!! #####
#####

# Anzahl der zu simulierenden Metaanalysen
Meta.N<-10000

# Anzahl der Studien pro Metaanalyse
trials <- 3
#trials <- 4
#trials <- 5
#trials <- 6

# IM WEITEREN PROGRAMM MÜSSEN KEINE WEITEREN PARAMETER EINGESTELLT WERDEN
#####
#####
# Bugs-Modell Datei erstellen: Random-Modell
# Varianz zwischen den Studien wird zusätzlich geschätzt.

random.model <- function(){

  #Simulation mit Uniform(0,100) als hyper a-priori Verteilung für tau
  for (a in A:B)

```



```

{
# Erwachsenen- oder Kinderdaten
# Daten Modell (Likelihood)
# Y[a]: observed ln OR in trial a:
Y.Uni[a] <- log( ev.trt[a] * (n.ctrl[a]-ev.ctrl[a]) / ( ev.ctrl[a] * (n.trt[a]-ev.trt[a]) ) )
VARlnOR.Uni[a] <- (1/ev.trt[a]) + (1/(n.trt[a]-ev.trt[a])) +
(1/ev.ctrl[a]) + (1/(n.ctrl[a]-ev.ctrl[a]))
PREClNOR.Uni[a] <- 1/VARlnOR.Uni[a]
Y.Uni[a] ~ dnorm(theta.Uni[a],PREClNOR.Uni[a])

# Parameter Model
theta.Uni[a] ~ dnorm(mu.Uni,Prec.theta.Uni)
}

# A-priori Verteilung
# mean(mu) entspricht pooled log OR
mu.Uni ~ dnorm(0,1.0E-6)

###Prior = Gleichverteilung
tau.Uni ~ dunif(1.0E-6, 100)
Prec.theta.Uni <- 1/(tau.Uni*tau.Uni)

# theta* Vorhersage für eine neue Studie
theta.new.Uni ~ dnorm(mu.Uni,Prec.theta.Uni)

#####
#Simulation mit Half-Normal(0,1) als hyper a-priori Verteilung für tau
for (a in A:B)
{
# Erwachsenen- oder Kinderdaten
# Daten Modell (Likelihood)
# Y[a]: observed ln OR in trial a:
Y.HN[a] <- log( ev.trt[a] * (n.ctrl[a]-ev.ctrl[a]) / ( ev.ctrl[a] * (n.trt[a]-ev.trt[a]) ) )
VARlnOR.HN[a] <- (1/ev.trt[a]) + (1/(n.trt[a]-ev.trt[a])) +
(1/ev.ctrl[a]) + (1/(n.ctrl[a]-ev.ctrl[a]))
PREClNOR.HN[a] <- 1/VARlnOR.HN[a]
Y.HN[a] ~ dnorm(theta.HN[a],PREClNOR.HN[a])

# Parameter Model
theta.HN[a] ~ dnorm(mu.HN,Prec.theta.HN)
}

# A-priori Verteilung
# mean(mu) entspricht pooled log OR
mu.HN ~ dnorm(0,1.0E-6)

### Prior = Half-Normal Verteilung
tau.HN ~ dnorm(0,1)%_I(0,)
Prec.theta.HN <- 1/(tau.HN*tau.HN)

# theta* Vorhersage für eine neue Studie
theta.new.HN ~ dnorm(mu.HN,Prec.theta.HN)

#####
#Simulation mit Inverse-Gamma(0.001,0.001) als hyper a-priori Verteilung für tau^2
for (a in A:B)
{
# Erwachsenen- oder Kinderdaten
# Daten Modell (Likelihood)
# Y[a]: observed ln OR in trial a:
Y.IG[a] <- log( ev.trt[a] * (n.ctrl[a]-ev.ctrl[a]) / ( ev.ctrl[a] * (n.trt[a]-ev.trt[a]) ) )
VARlnOR.IG[a] <- (1/ev.trt[a]) + (1/(n.trt[a]-ev.trt[a])) +
(1/ev.ctrl[a]) + (1/(n.ctrl[a]-ev.ctrl[a]))
PREClNOR.IG[a] <- 1/VARlnOR.IG[a]
Y.IG[a] ~ dnorm(theta.IG[a],PREClNOR.IG[a])

# Parameter Model
theta.IG[a] ~ dnorm(mu.IG,Prec.theta.IG)
}

# A-priori Verteilung
# mean(mu) entspricht pooled log OR
mu.IG ~ dnorm(0,1.0E-6)

###Prior = Inverse-Gamma-Verteilung
Prec.theta.IG ~ dgamma(0.001,0.001)
# tau ist die between-trial variation of study-specific effects
tau.IG <- 1/(sqrt(Prec.theta.IG))

# theta* Vorhersage für eine neue Studie
theta.new.IG ~ dnorm(mu.IG,Prec.theta.IG)

#####
#Simulation mit Half-Cauchy(0,1) als hyper a-priori Verteilung für tau
for (a in A:B)
{
# Erwachsenen- oder Kinderdaten
# Daten Modell (Likelihood)
# Y[a]: observed ln OR in trial a:
Y.HC[a] <- log( ev.trt[a] * (n.ctrl[a]-ev.ctrl[a]) / ( ev.ctrl[a] * (n.trt[a]-ev.trt[a]) ) )
VARlnOR.HC[a] <- (1/ev.trt[a]) + (1/(n.trt[a]-ev.trt[a])) +
(1/ev.ctrl[a]) + (1/(n.ctrl[a]-ev.ctrl[a]))
PREClNOR.HC[a] <- 1/VARlnOR.HC[a]
Y.HC[a] ~ dnorm(theta.HC[a],PREClNOR.HC[a])
}

```

```

# Parameter Model
theta.HC[a] ~ dnorm(mu.HC, Prec.theta.HC)
}

# A-priori Verteilung
# mean(mu) entspricht pooled log OR
mu.HC ~ dnorm(0,1.0E-6)

###Prior = Half-Cauchy Verteilung
### half-cauchy(sigma2/sigma1)=abs(N(0,sigma1^2)/N(0,sigma2^2)
### siehe Paper S. 4/5:
### "Marley JK, Wand MP - (2010) - Non-Standard Semiparametric Regression via BRugs"

prior.scale.HC <- 1
prec.prior.scale.HC <- 1/(prior.scale.HC*prior.scale.HC)
norm1.HC ~ dnorm(0,1)
norm2.HC ~ dnorm(0,prec.prior.scale.HC)
Prec.theta.HC <- pow(norm1.HC/norm2.HC,2)
tau.HC <- 1/(sqrt(Prec.theta.HC))

# theta* Vorhersage für eine neue Studie
theta.new.HC ~ dnorm(mu.HC, Prec.theta.HC)

}
#####

# Modell als Textdatei schreiben
write.model(random.model, con = "Modell-Random.txt")

#####

# Bugs-Modell Datei erstellen: Referenz-Modell, Varianz zwischen den Studien ist Null

# Modell als R-Funktion (WinBUGS Syntax)
referenz.model <- function(){
  for (a in A:B)
  {
    # Data
    # Daten Modell (Likelihood)
    # Y[a]: observed ln OR in trial a:
    Y[a] <- log( ev.trt[a] * (n.ctrl[a]-ev.ctrl[a]) /
      ( ev.ctrl[a] * (n.trt[a]-ev.trt[a]) ) )
    VARlnOR[a] <- (1/ev.trt[a]) + (1/(n.trt[a]-ev.trt[a])) +
      (1/ev.ctrl[a]) + (1/(n.ctrl[a]-ev.ctrl[a]))
    PREClnOR[a] <- 1/VARlnOR[a]
    Y[a] ~ dnorm(mu, PREClnOR[a])
  }

  # A-priori Verteilung
  mu ~ dnorm(0,1.0E-6)

  # theta* Vorhersage für eine neue Studie
  theta.new <- mu
}

# Modell als Textdatei schreiben
write.model(referenz.model, con = "Referenz.txt")

#####
#MACRO zur Generierung des Datensatzes und Berechnung von Neff
run.test <- defmacro(simulationen,expr={

# Erzeugung der Daten für eine Metaanalyse
Datenliste <- list()
for(ijk in 1:trials){

# Erzeugung von Kontingenztafeln zur Bestimmung der wahren Ereigniswahrscheinlichkeiten
# für jede einzelne Studie

logOR=rnorm(1,mean=mu_0,sd=tau_0) # Generierung des studienspezifischen wahren logOR
OR <- exp(logOR) #Wahre Odds-Ratio

N <- 10000000 # Gesamtfallzahl der Kontingenztafel, die bezgl. des wahren studienspezifischen
# logOR erzeugt wird.
# N ist so groß gewählt, damit wir die wahren Ereigniswahrscheinlichkeit genau
# genug bestimmen können.

P <- pTC
if(P == 0){P <- 0.5/N}
# cat(P, "\n")
nT <- round(N/2)
nC <- N - nT
eTC <- round(P * N)
# Generierung einer Kontingenztafel zum wahren studienspezifischen logOR
Kontingenztafel1 <- Kontingenztafel(nT,nC,eTC,OR)

# Berechnung der wahren studienspezifischen Ereigniswahrscheinlichkeiten
tT <- Kontingenztafel1$EreignisT/Kontingenztafel1$Treat
tC <- Kontingenztafel1$EreignisK/Kontingenztafel1$Kontrolle

N0=floor(n.trials/2)
N1=ceiling(n.trials/2)

# Zufällig Ziehung der Ereignisanzahlen für die einzelnen Studie

```

```

# aus den berechneten wahren studienspezifischen Ereigniswahrscheinlichkeiten
# auf Grundlage der gewählten Fallzahl pro Studie.
eT = rbinom(1, size=N1, prob=tT)
eC = rbinom(1, size=N0, prob=tC)
daten1 <- data.frame(eT = eT, eC = eC, nT = N1, nC = N0)
Datenliste[[ijk]] <- daten1
}
Daten <- do.call(rbind, Datenliste)

data <- list (A=1, B=trials,
            n.trt = c(Daten$nT),
            ev.trt = c(Daten$eT),
            n.ctrl = c(Daten$nC),
            ev.ctrl = c(Daten$eC))

#####
# Berechnung der "wahren" Precision Weight Effective Sample Size N.true!
#####

N.hist=trials*n.trials
Daten$var.within <- 1/Daten$eT + 1/(Daten$nT-Daten$eT) + 1/Daten$eC + 1/(Daten$nC-Daten$eC)
N.true <- (1/sum((Daten$var.within)^(-1)) / ((1 / sum( (Daten$var.within+tau_0^2)^(-1)) ) + tau_0^2)) * N.hist

#####
# STARTEN DER MCMC-SAMPLINGS
# Volles Modell (Random-Effect-Modell)
#
# Startwerte der Markovketten: inits1 for Markov Chain 1
#           inits2 for Markov Chain 2
# Je nach Priorverteilung muss tau oder prec.theta Startwert übergeben werden
inits1 <- list(mu.HN=0, tau.HN=1, mu.Uni=0, tau.Uni=1, mu.IG=0, prec.theta.IG=0.2, mu.HC=0)
inits2 <- list(mu.HN=-10, tau.HN=0.5, mu.Uni=-10, tau.Uni=0.5, mu.IG=-10, prec.theta.IG=0.5, mu.HC=-10)

inits <- list(inits1, inits2)

# Ausführen des Modells
meta.sim <- bugs(
  program="openbugs",
  #program="winbugs",
  data,inits,
  model.file="Modell-Random.txt",
  parameters= c("tau.Uni","theta.new.Uni",
               "tau.HC", "theta.new.HC",
               "tau.HN", "theta.new.HN",
               "tau.IG", "theta.new.IG"),
  n.chains = 2,
  n.iter=110000,
  n.burnin=10000,
  n.thin=10,
  bugs.directory=OpenBUGSdirectory,
  #bugs.directory=WinBUGSdirectory,
  codaPkg=FALSE,
  debug=TRUE)

#####
# Referenz: Fixed effect model. Annahme: die Interstudienstandardabweichung tau=0
#####

# Startwerte der Markovketten
inits<-list(list(mu=0),list(mu=-10))

# Ausführen des Modells
metaref.sim <- bugs(
  program="openbugs",
  #program="winbugs",
  data,inits,
  model.file="Referenz.txt",
  parameters=c("mu","theta.new"),
  n.chains = 2,
  n.iter=110000,
  n.burnin=10000,
  n.thin=10,
  bugs.directory=OpenBUGSdirectory,
  #bugs.directory=WinBUGSdirectory,
  debug=TRUE)

#print(metaref.sim, digits=4)

#####
random<-meta.sim$sims.array
random_m<-meta.sim$sims.matrix
#random_m[1:10,]
ref<-metaref.sim$sims.array
ref_m<-metaref.sim$sims.matrix

#####
# Trimmen des Random-Modells bezgl. theta.new #

theta.Uni.percentil.low <- quantile(random[, "theta.new.Uni"], c(.001))
theta.Uni.percentil.up <- quantile(random[, "theta.new.Uni"], c(.999))
trimmed.random.Uni <- random_m[ random_m[, "theta.new.Uni"] > theta.Uni.percentil.low
& random_m[, "theta.new.Uni"] < theta.Uni.percentil.up, ]

theta.IG.percentil.low <- quantile(random[, "theta.new.IG"], c(.001))
theta.IG.percentil.up <- quantile(random[, "theta.new.IG"], c(.999))

```

```

trimmed.random.IG <- random_m[ random_m[ ,"theta.new.IG"] > theta.IG.percentil.low
& random_m[ ,"theta.new.IG"] < theta.IG.percentil.up, ]

theta.HN.percentil.low <- quantile(random[, ,"theta.new.HN"], c(.001))
theta.HN.percentil.up <- quantile(random[, ,"theta.new.HN"], c(.999))
trimmed.random.HN <- random_m[ random_m[ ,"theta.new.HN"] > theta.HN.percentil.low
& random_m[ ,"theta.new.HN"] < theta.HN.percentil.up, ]

theta.HC.percentil.low <- quantile(random[, ,"theta.new.HC"], c(.001))
theta.HC.percentil.up <- quantile(random[, ,"theta.new.HC"], c(.999))
trimmed.random.HC <- random_m[ random_m[ ,"theta.new.HC"] > theta.HC.percentil.low
& random_m[ ,"theta.new.HC"] < theta.HC.percentil.up, ]

#####
#####
#
#Übersetzung der a-posteriori prädiktiven Präzisionen:
# Berechnung des Gewichtungsfaktors und Neff
#
#####
#Unverändertes Modell
#####

n.gesamt<-trials*n.trials

omega.Uni <- (sd(ref_m[,"theta.new"]))^2/((sd(random_m[,"theta.new.Uni"]))^2)
Neff.Uni <- omega.Uni*n.gesamt
prozent.Uni<- 100*omega.Uni

omega.HC <- (sd(ref_m[,"theta.new"]))^2/((sd(random_m[,"theta.new.HC"]))^2)
Neff.HC <- omega.HC*n.gesamt
prozent.HC<- 100*omega.HC

omega.HN <- (sd(ref_m[,"theta.new"]))^2/((sd(random_m[,"theta.new.HN"]))^2)
Neff.HN <- omega.HN*n.gesamt
prozent.HN<- 100*omega.HN

omega.IG <- (sd(ref_m[,"theta.new"]))^2/((sd(random_m[,"theta.new.IG"]))^2)
Neff.IG <- omega.IG*n.gesamt
prozent.IG<- 100*omega.IG

#####
#Getrimmtes Random Modell
#####
trimmed.omega.Uni <- (sd(ref_m[,"theta.new"]))^2/((sd(trimmed.random.Uni[,"theta.new.Uni"]))^2)
trimmed.Neff.Uni <- trimmed.omega.Uni*n.gesamt
trimmed.prozent.Uni<- 100*trimmed.omega.Uni

trimmed.omega.HC <- (sd(ref_m[,"theta.new"]))^2/((sd(trimmed.random.HC[,"theta.new.HC"]))^2)
trimmed.Neff.HC <- trimmed.omega.HC*n.gesamt
trimmed.prozent.HC<- 100*trimmed.omega.HC

trimmed.omega.HN <- (sd(ref_m[,"theta.new"]))^2/((sd(trimmed.random.HN[,"theta.new.HN"]))^2)
trimmed.Neff.HN <- trimmed.omega.HN*n.gesamt
trimmed.prozent.HN<- 100*trimmed.omega.HN

trimmed.omega.IG <- (sd(ref_m[,"theta.new"]))^2/((sd(trimmed.random.IG[,"theta.new.IG"]))^2)
trimmed.Neff.IG <- trimmed.omega.IG*n.gesamt
trimmed.prozent.IG<- 100*trimmed.omega.IG

#rm(i)
})

#####
#Aufrufbefehl des Macros
#run.test(Simulationen)

#####
# Ergebnisse der Simulationen erzeugen.

# Anlegen von Vektoren in denen die jeweiligen Ergebnisse gespeichert werden.
Neff.untrimmed.Uni <- numeric(length=Meta.N)
Neff.untrimmed.HN <- numeric(length=Meta.N)
Neff.untrimmed.HC <- numeric(length=Meta.N)
Neff.untrimmed.IG <- numeric(length=Meta.N)

omega.untrimmed.Uni <- numeric(length=Meta.N)
omega.untrimmed.HN <- numeric(length=Meta.N)
omega.untrimmed.HC <- numeric(length=Meta.N)
omega.untrimmed.IG <- numeric(length=Meta.N)

theta.new.untrimmed.Uni <- numeric(length=Meta.N)
theta.new.untrimmed.HN <- numeric(length=Meta.N)
theta.new.untrimmed.HC <- numeric(length=Meta.N)
theta.new.untrimmed.IG <- numeric(length=Meta.N)

sqrt.V.tau.new.untrimmed.Uni <- numeric(length=Meta.N)
sqrt.V.tau.new.untrimmed.HN <- numeric(length=Meta.N)
sqrt.V.tau.new.untrimmed.HC <- numeric(length=Meta.N)
sqrt.V.tau.new.untrimmed.IG <- numeric(length=Meta.N)

tau.untrimmed.Uni <- numeric(length=Meta.N)
tau.untrimmed.HN <- numeric(length=Meta.N)
tau.untrimmed.HC <- numeric(length=Meta.N)
tau.untrimmed.IG <- numeric(length=Meta.N)

```

```

Neff.trimmed.Uni <- numeric(length=Meta.N)
Neff.trimmed.HN <- numeric(length=Meta.N)
Neff.trimmed.HC <- numeric(length=Meta.N)
Neff.trimmed.IG <- numeric(length=Meta.N)

omega.trimmed.Uni <- numeric(length=Meta.N)
omega.trimmed.HN <- numeric(length=Meta.N)
omega.trimmed.HC <- numeric(length=Meta.N)
omega.trimmed.IG <- numeric(length=Meta.N)

theta.new.trimmed.Uni <- numeric(length=Meta.N)
theta.new.trimmed.HN <- numeric(length=Meta.N)
theta.new.trimmed.HC <- numeric(length=Meta.N)
theta.new.trimmed.IG <- numeric(length=Meta.N)

sqrt.V.tau.new.trimmed.Uni <- numeric(length=Meta.N)
sqrt.V.tau.new.trimmed.HN <- numeric(length=Meta.N)
sqrt.V.tau.new.trimmed.HC <- numeric(length=Meta.N)
sqrt.V.tau.new.trimmed.IG <- numeric(length=Meta.N)

tau.trimmed.Uni <- numeric(length=Meta.N)
tau.trimmed.HN <- numeric(length=Meta.N)
tau.trimmed.HC <- numeric(length=Meta.N)
tau.trimmed.IG <- numeric(length=Meta.N)

N.truesize <- numeric(length=Meta.N)

# Ausführen der Simulationen
for(k in 1:Meta.N){
  run.test(Simulationen[[k]]) # Aufruf des Simulationsmakros

  Neff.untrimmed.Uni[k] <- Neff.Uni
  Neff.untrimmed.HN[k] <- Neff.HN
  Neff.untrimmed.HC[k] <- Neff.HC
  Neff.untrimmed.IG[k] <- Neff.IG

  omega.untrimmed.Uni[k] <- omega.Uni
  omega.untrimmed.HN[k] <- omega.HN
  omega.untrimmed.HC[k] <- omega.HC
  omega.untrimmed.IG[k] <- omega.IG

  theta.new.untrimmed.Uni[k] <- mean(random[, "theta.new.Uni"])
  theta.new.untrimmed.HN[k] <- mean(random[, "theta.new.HN"])
  theta.new.untrimmed.HC[k] <- mean(random[, "theta.new.HC"])
  theta.new.untrimmed.IG[k] <- mean(random[, "theta.new.IG"])

  sqrt.V.tau.new.untrimmed.Uni[k] <- sd(random_m[, "theta.new.Uni"])
  sqrt.V.tau.new.untrimmed.HN[k] <- sd(random_m[, "theta.new.HN"])
  sqrt.V.tau.new.untrimmed.HC[k] <- sd(random_m[, "theta.new.HC"])
  sqrt.V.tau.new.untrimmed.IG[k] <- sd(random_m[, "theta.new.IG"])

  tau.untrimmed.Uni[k] <- mean(random[, "tau.Uni"])
  tau.untrimmed.HN[k] <- mean(random[, "tau.HN"])
  tau.untrimmed.HC[k] <- mean(random[, "tau.HC"])
  tau.untrimmed.IG[k] <- mean(random[, "tau.IG"])

  Neff.trimmed.Uni[k] <- trimmed.Neff.Uni
  Neff.trimmed.HN[k] <- trimmed.Neff.HN
  Neff.trimmed.HC[k] <- trimmed.Neff.HC
  Neff.trimmed.IG[k] <- trimmed.Neff.IG

  omega.trimmed.Uni[k] <- trimmed.omega.Uni
  omega.trimmed.HN[k] <- trimmed.omega.HN
  omega.trimmed.HC[k] <- trimmed.omega.HC
  omega.trimmed.IG[k] <- trimmed.omega.IG

  theta.new.trimmed.Uni[k] <- mean(trimmed.random.Uni[, "theta.new.Uni"])
  theta.new.trimmed.HN[k] <- mean(trimmed.random.HN[, "theta.new.HN"])
  theta.new.trimmed.HC[k] <- mean(trimmed.random.HC[, "theta.new.HC"])
  theta.new.trimmed.IG[k] <- mean(trimmed.random.IG[, "theta.new.IG"])

  sqrt.V.tau.new.trimmed.Uni[k] <- sd(trimmed.random.Uni[, "theta.new.Uni"])
  sqrt.V.tau.new.trimmed.HN[k] <- sd(trimmed.random.HN[, "theta.new.HN"])
  sqrt.V.tau.new.trimmed.HC[k] <- sd(trimmed.random.HC[, "theta.new.HC"])
  sqrt.V.tau.new.trimmed.IG[k] <- sd(trimmed.random.IG[, "theta.new.IG"])

  tau.trimmed.Uni[k] <- mean(trimmed.random.Uni[, "tau.Uni"])
  tau.trimmed.HN[k] <- mean(trimmed.random.HN[, "tau.HN"])
  tau.trimmed.HC[k] <- mean(trimmed.random.HC[, "tau.HC"])
  tau.trimmed.IG[k] <- mean(trimmed.random.IG[, "tau.IG"])

  N.truesize[k] <- N.true
  rm(meta.sim, metaref.sim, random, random_m, ref, ref_m)
}

#####
# Datensatz mit den Simulationsergebnissen erzeugen
# Ergebnisse der einzelnen hyper a-priori Verteilungen stehen untereinander
# Gruppierungsvariable heißt Prior

Simulation.Resultat <- data.frame(
  Neff.untrimmed=NA,
  omega.untrimmed=NA,
  theta.new.untrimmed=NA,
  sqrt.V.tau.new.untrimmed=NA,

```

```

tau.untrimmed=NA,
Neff.trimmed=NA,
omega.trimmed=NA,
theta.new.trimmed=NA,
sqrt.V.tau.new.trimmed=NA,
tau.trimmed=NA,
prior=NA,
N.true=NA)

#Simulation.Resultat
Simulation.Resultat[1:Meta.N,"Neff.untrimmed"] <- Neff.untrimmed.Uni
Simulation.Resultat[(Meta.N+1):(2*Meta.N),"Neff.untrimmed"] <- Neff.untrimmed.HN
Simulation.Resultat[(2*Meta.N+1):(3*Meta.N),"Neff.untrimmed"] <- Neff.untrimmed.HC
Simulation.Resultat[(3*Meta.N+1):(4*Meta.N),"Neff.untrimmed"] <- Neff.untrimmed.IG

Simulation.Resultat[1:Meta.N,"omega.untrimmed"] <- omega.untrimmed.Uni
Simulation.Resultat[(Meta.N+1):(2*Meta.N),"omega.untrimmed"] <- omega.untrimmed.HN
Simulation.Resultat[(2*Meta.N+1):(3*Meta.N),"omega.untrimmed"] <- omega.untrimmed.HC
Simulation.Resultat[(3*Meta.N+1):(4*Meta.N),"omega.untrimmed"] <- omega.untrimmed.IG

Simulation.Resultat[1:Meta.N,"theta.new.untrimmed"] <-theta.new.untrimmed.Uni
Simulation.Resultat[(Meta.N+1):(2*Meta.N),"theta.new.untrimmed"] <-theta.new.untrimmed.HN
Simulation.Resultat[(2*Meta.N+1):(3*Meta.N),"theta.new.untrimmed"] <-theta.new.untrimmed.HC
Simulation.Resultat[(3*Meta.N+1):(4*Meta.N),"theta.new.untrimmed"] <-theta.new.untrimmed.IG

Simulation.Resultat[1:Meta.N,"sqrt.V.tau.new.untrimmed"] <-sqrt.V.tau.new.untrimmed.Uni
Simulation.Resultat[(Meta.N+1):(2*Meta.N),"sqrt.V.tau.new.untrimmed"] <-sqrt.V.tau.new.untrimmed.HN
Simulation.Resultat[(2*Meta.N+1):(3*Meta.N),"sqrt.V.tau.new.untrimmed"] <-sqrt.V.tau.new.untrimmed.HC
Simulation.Resultat[(3*Meta.N+1):(4*Meta.N),"sqrt.V.tau.new.untrimmed"] <-sqrt.V.tau.new.untrimmed.IG

Simulation.Resultat[1:Meta.N,"tau.untrimmed"] <-tau.untrimmed.Uni
Simulation.Resultat[(Meta.N+1):(2*Meta.N),"tau.untrimmed"] <-tau.untrimmed.HN
Simulation.Resultat[(2*Meta.N+1):(3*Meta.N),"tau.untrimmed"] <-tau.untrimmed.HC
Simulation.Resultat[(3*Meta.N+1):(4*Meta.N),"tau.untrimmed"] <-tau.untrimmed.IG

Simulation.Resultat[1:Meta.N,"Neff.trimmed"] <- Neff.trimmed.Uni
Simulation.Resultat[(Meta.N+1):(2*Meta.N),"Neff.trimmed"] <- Neff.trimmed.HN
Simulation.Resultat[(2*Meta.N+1):(3*Meta.N),"Neff.trimmed"] <- Neff.trimmed.HC
Simulation.Resultat[(3*Meta.N+1):(4*Meta.N),"Neff.trimmed"] <- Neff.trimmed.IG

Simulation.Resultat[1:Meta.N,"omega.trimmed"] <- omega.trimmed.Uni
Simulation.Resultat[(Meta.N+1):(2*Meta.N),"omega.trimmed"] <- omega.trimmed.HN
Simulation.Resultat[(2*Meta.N+1):(3*Meta.N),"omega.trimmed"] <- omega.trimmed.HC
Simulation.Resultat[(3*Meta.N+1):(4*Meta.N),"omega.trimmed"] <- omega.trimmed.IG

Simulation.Resultat[1:Meta.N,"theta.new.trimmed"] <-theta.new.trimmed.Uni
Simulation.Resultat[(Meta.N+1):(2*Meta.N),"theta.new.trimmed"] <-theta.new.trimmed.HN
Simulation.Resultat[(2*Meta.N+1):(3*Meta.N),"theta.new.trimmed"] <-theta.new.trimmed.HC
Simulation.Resultat[(3*Meta.N+1):(4*Meta.N),"theta.new.trimmed"] <-theta.new.trimmed.IG

Simulation.Resultat[1:Meta.N,"sqrt.V.tau.new.trimmed"] <-sqrt.V.tau.new.trimmed.Uni
Simulation.Resultat[(Meta.N+1):(2*Meta.N),"sqrt.V.tau.new.trimmed"] <-sqrt.V.tau.new.trimmed.HN
Simulation.Resultat[(2*Meta.N+1):(3*Meta.N),"sqrt.V.tau.new.trimmed"] <-sqrt.V.tau.new.trimmed.HC
Simulation.Resultat[(3*Meta.N+1):(4*Meta.N),"sqrt.V.tau.new.trimmed"] <-sqrt.V.tau.new.trimmed.IG

Simulation.Resultat[1:Meta.N,"tau.trimmed"] <-tau.trimmed.Uni
Simulation.Resultat[(Meta.N+1):(2*Meta.N),"tau.trimmed"] <-tau.trimmed.HN
Simulation.Resultat[(2*Meta.N+1):(3*Meta.N),"tau.trimmed"] <-tau.trimmed.HC
Simulation.Resultat[(3*Meta.N+1):(4*Meta.N),"tau.trimmed"] <-tau.trimmed.IG

Simulation.Resultat[1:Meta.N,"prior"] <- "Uniform"
Simulation.Resultat[(Meta.N+1):(2*Meta.N),"prior"] <- "Half-Normal"
Simulation.Resultat[(2*Meta.N+1):(3*Meta.N),"prior"] <- "Half-Cauchy"
Simulation.Resultat[(3*Meta.N+1):(4*Meta.N),"prior"] <- "Inverse-Gamma"

Simulation.Resultat[1:Meta.N,"N.true"] <- N.truesize
Simulation.Resultat[(Meta.N+1):(2*Meta.N),"N.true"] <- N.truesize
Simulation.Resultat[(2*Meta.N+1):(3*Meta.N),"N.true"] <- N.truesize
Simulation.Resultat[(3*Meta.N+1):(4*Meta.N),"N.true"] <- N.truesize

#Simulation.Resultat

#####
#####
# POWERANALYSE #####
#####

# Powerfunktion erstellen zum Einbringen von Prior-Information
#####
# Neue Studie mit Normal-Normal Modell
# Wir nehmen an, dass die Studien nur um das wahre gepoolte logOR schwanken.

# Funktion zur Simulation der Power unter Verwendung informativer a-priori Parameter
Power.Berechnung <- function(ntotal.eff)
{ Var.logORnew <- Sd.logORnew^2
  power.status <- NA
  sim.power<- NA
  ntotal.eff <- round(N_ref - Neff)
  if(ntotal.eff>=0)
  {
    nC <- floor(ntotal.eff/2)

```

```

nT <- ceiling(ntotal.eff/2)

power.data <- data.frame(eT = rbinom(lauf, size=nT, prob=pT),
  eC = rbinom(lauf, size=nC, prob=pC))
power.data$logOR <- log(power.data$eT*(nC-power.data$eC)/power.data$eC/(nT-power.data$eT))
power.data$var.logOR <- 1/power.data$eT + 1/(nT-power.data$eT) + 1/power.data$eC + 1/(nC-power.data$eC)

power.data$logOR.UCL <- ( (1/(Var.logORnew)) * ( E.logOR) + (1/power.data$var.logOR) * power.data$logOR ) /
  ( (1/(Var.logORnew)) + (1/power.data$var.logOR) ) +
  qnorm(0.975) * sqrt( 1 / ( (1/(Var.logORnew)) + (1/power.data$var.logOR) ) ) )

sim.power <- mean(power.data$logOR.UCL < 0, na.rm=TRUE)

if(!is.na(power.data$logOR.UCL[1]))
{
  if(power.data$logOR.UCL[1]<0){power.status <- 1}
  if(power.data$logOR.UCL[1]>=0){power.status <- 0}
}
}
return(list(sim.power, power.status))
}

# POWER SIMULATIONEN MIT REDUZIERTER FALLZAHL

lauf <- 100000 # Anzahl der Simulationen der Ergebnisse der neuen Studie
#####
# Prior Information:

power.untrimmed <- NA
powerstatus.untrimmed <- NA
for(lmo in 1:(4*Meta.N)){
  E.logOR <- Simulation.Resultat[ , "theta.new.untrimmed"][lmo]
  Sd.logORnew <- Simulation.Resultat[ , "sqrt.V.tau.new.untrimmed"][lmo]
  Neff <- Simulation.Resultat[ , "Neff.untrimmed"][lmo]

  power.list <- Power.Berechnung(ntotal.eff)
  power.untrimmed[lmo] <- power.list[[1]]
  powerstatus.untrimmed[lmo] <- power.list[[2]]
  rm(E.logOR, Sd.logORnew, Neff)
}
Simulation.Resultat[1:(4*Meta.N), "power.untrimmed"] <- power.untrimmed
Simulation.Resultat[1:(4*Meta.N), "powerstatus.untrimmed"] <- powerstatus.untrimmed

power.true <- NA
powerstatus.true <- NA
for(lmo in 1:(4*Meta.N)){
  E.logOR <- Simulation.Resultat[ , "theta.new.untrimmed"][lmo]
  Sd.logORnew <- Simulation.Resultat[ , "sqrt.V.tau.new.untrimmed"][lmo]
  Neff <- Simulation.Resultat[ , "N.true"][lmo]
  power.list <- Power.Berechnung(ntotal.eff)
  power.true[lmo] <- power.list[[1]]
  powerstatus.true[lmo] <- power.list[[2]]
  rm(E.logOR, Sd.logORnew, Neff)
}
Simulation.Resultat[1:(4*Meta.N), "power.true"] <- power.true
Simulation.Resultat[1:(4*Meta.N), "powerstatus.true"] <- powerstatus.true

power.trimmed <- NA
powerstatus.trimmed <- NA
for(lmo in 1:(4*Meta.N)){
  E.logOR <- Simulation.Resultat[ , "theta.new.trimmed"][lmo]
  Sd.logORnew <- Simulation.Resultat[ , "sqrt.V.tau.new.trimmed"][lmo]
  Neff <- Simulation.Resultat[ , "Neff.trimmed"][lmo]

  power.list <- Power.Berechnung(ntotal.eff)
  power.trimmed[lmo] <- power.list[[1]]
  powerstatus.trimmed[lmo] <- power.list[[2]]
  rm(E.logOR, Sd.logORnew, Neff)
}
Simulation.Resultat[1:(4*Meta.N), "power.trimmed"] <- power.trimmed
Simulation.Resultat[1:(4*Meta.N), "powerstatus.trimmed"] <- powerstatus.trimmed
#Simulation.Resultat

# Deskriptive Statistiken über alle Variablen getrennt nach a-priori Verteilung
# lapply(split(Simulation.Resultat, Simulation.Resultat[, "prior"]), summary)

#####
#####
# Export der Simulationsresultate #####
#####
#####

library(foreign)
# Export als *.csv-Datei: write.csv2 benutzt Komma für Dezimalpunkte und Semikolon als Separator
write.csv2(Simulation.Resultat, outputcsvname)

```

Literatur

- [1] ALSMEYER, G.: *Mathematische Statistik*. 3. Aufl. Skripten zur Mathematischen Statistik 36. Institut für Mathematische Statistik der Westfälischen Wilhelms-Universität Münster, 2006.
- [2] ALTMAN, D. G.: Discussion of Bayesian approaches to randomised trials by Spiegelhalter et. al. In: *Journal of the Royal Statistical Society*, Bd. 157 (1994), S. 387–416.
- [3] ALTMAN, D. G. und BLAND, J. M.: Statistics notes. Treatment allocation in controlled trials: why randomise? In: *BMJ*, Bd. 318(7192) (1999), S. 1209.
- [4] ANDREWS, D. F.: *Robust estimates of location: survey and advances*. Princeton University Press, Princeton, N.J., 1972, S. 373.
- [5] ARENDS, L. R.; HUNINK, M. G. M. und STIJNEN, T.: Meta-analysis of summary survival curve data. In: *Statistics in Medicine*, Bd. 27(22) (2008), S. 4381–4396.
- [6] BAUR, X. und BARBINOVA, L.: Messung von exhalierem Stickstoffmonoxid: Klinischer Einsatz bei Atemwegserkrankungen. In: *Dtsch Arztebl Int*, Bd. 104(12) (2007),
- [7] BAYES, T.: An essay towards solving a problem in the doctrine of chances. In: *Phil. Trans. of the Royal Soc. of London*, Bd. 53 (1763), S. 370–418.
- [8] BERRY, S.; CARLIN, B.; LEE, J. und MÜLLER, P.: *Bayesian Adaptive Methods for Clinical Trials*. Chapman & Hall/Crc Biostatistics Series. CRC Press, 2010.
- [9] BLAND, M. J. und ALTMAN, D. G.: Statistics notes: Bayesians and frequentists. In: *BMJ*, Bd. 317(7166) (1998), S. 1151–1160.
- [10] BOLSTAD, W.: *Understanding computational Bayesian statistics*. Wiley series in computational statistics. Wiley, 2010.
- [11] CHEN, M.-H. und IBRAHIM, J. G.: The Relationship Between the Power Prior and Hierarchical Models. In: *Bayesian Analysis*, Bd. 1(3) (2006), S. 551–574.
- [12] DANIELS, M. J.: A Prior for the Variance in Hierarchical Models. In: *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, Bd. 27(3) (1999), S. 567–578.
- [13] DERSIMONIAN, R. und LAIRD, N.: Meta-analysis in clinical trials. In: *Controlled Clinical Trials*, Bd. 7(3) (1986), S. 177–188.
- [14] DREW, J.; EVANS, D. und GLEN, A.: *Computational Probability: Algorithms and Applications in the Mathematical Sciences*. Bd. 117. International series in operations research & management science. Springer, 2008.
- [15] ELSTRODT, J.: *Maß- und Integrationstheorie*. Springer-Lehrbuch. Springer London, Limited, 2011.

- [16] ERKENS, P. M. und PRINS, M. H.: Fixed dose subcutaneous low molecular weight heparins versus adjusted dose unfractionated heparin for venous thromboembolism. In: *Cochrane Database Syst Rev*, Bd. (9) (2010), S. CD001100.
- [17] EUROPEAN MEDICINES AGENCY (EMA): *Committee for Medicinal Products for Human Use (CHMP): Guideline on Clinical Trials in Small Populations*. London, 2005.
- [18] EUROPEAN MEDICINES AGENCY (EMA): *Committee for Medicinal Products for Human Use (CHMP): Reflection Paper on Methodological Issues in Confirmatory Clinical Trials with Flexible Design and Analysis Plan*. London, 2006.
- [19] EUROPEAN MEDICINES AGENCY (EMA): *ICH Topic E10: Choice of Control Group in Clinical Trials - Note for Guidance in choice of control group in clinical trials*. London, 2001.
- [20] GAMERMAN, D. und LOPES, H. F.: *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman & Hall/CRC Texts in Statistical Science, 2006.
- [21] GELMAN, A.: Prior distributions for variance parameters in hierarchical models. In: *Bayesian Analysis*, Bd. 1 (2006), S. 1–19.
- [22] GELMAN, A.; CARLIN, J. B.; STERN, H. S. und RUBIN, D. B.: *Bayesian Data Analysis, Second Edition*. 2nd. Chapman & Hall/CRC Texts in Statistical Science, 2003.
- [23] GERSS, J.; MALE, C.; BENDA, N. und KÖPCKE, W.: Ein Bayes-Ansatz zur Planung einer pädiatrischen klinischen Studie, in deren Auswertung die vorliegenden Daten aus Studien an Erwachsenen eingebracht werden. In: *Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie: Abstractband der 55. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS) e.V. ; 55.2010*. Hrsg. von SCHMÜCKER, P.; ELLSÄSSER, K.-H. und HAYNA, S. Mannheim: Dietzenbach : Antares-Computer-Verlag, 2010, S. 500–501.
- [24] GERSS, J.; MALE, C.; KOCH, R.; FALDUM, A.; KÖPCKE, W. und SCHMIDT, R.: Bayesian sample size estimation in a clinical trial utilizing historical information from different populations in the final analysis. submitted.
- [25] GOODMAN, S. N. und SLADKY, J. T.: A Bayesian approach to randomized controlled trials in children utilizing information from adults: the case of Guillain-Barre. In: *Clinical Trials*, Bd. 2(4) (2005), S. 305–310.
- [26] GOTTSCHALK, M.; DANNE, T.; VLAJNIC, A. und CARA, J. F.: Glimpiride versus metformin as monotherapy in pediatric patients with type 2 diabetes: a randomized, single-blind comparative study. In: *Diabetes Care*, Bd. 30(4) (2007), S. 790–794.
- [27] GREEN, S. und HIGGINS, J. P.: *Cochrane Handbook for Systematic Reviews of Interventions*. John Wiley & Sons, Ltd, 2008.

- [28] GREENHOUSE, J. B. und SELTMAN, H.: Using prior distributions to synthesize historical evidence: comments on the Goodman–Sladky case study of IVIg in Guillain–Barré syndrome. In: *Clinical Trials*, Bd. 2(4) (2005), S. 311–318.
- [29] GREENLAND, S.: Probability logic and probabilistic induction. In: *Epidemiology (Cambridge, Mass.)* Bd. 9(3) (1998), S. 322–332.
- [30] GSTEIGER, S.; NEUENSCHWANDER, B.; MERCIER, F. und SCHMIDLI, H.: Using historical control information for the design and analysis of clinical trials with overdispersed count data. In: *Statistics in Medicine*, Bd. 32(21) (2013), S. 3609–3622.
- [31] HASTINGS, W. K.: Monte Carlo sampling methods using Markov chains and their applications. In: *Biometrika*, Bd. 57(1) (1970), S. 97–109.
- [32] HIGGINS, J. P. T. und WHITEHEAD, A.: Borrowing strength from external trials in a meta-analysis. In: *Statistics in Medicine*, Bd. 15(24) (1996), S. 2733–2749.
- [33] IBRAHIM, J. G. und CHEN, M. H.: Power Prior Distributions for Regression Models. In: *Statistical Science*, Bd. 15(1) (2000), S. 46–60.
- [34] IRONY, T.; SINGPURWALLA, N. D.; COX, D. R.; DAWID, A. P.; GHOSH, J. K.; LINDLEY, D. und BERNARDO, J. M.: Non-informative priors do not exist: A dialogue with José M. Bernardo including discussion. In: *Journal of Statistical Planning and Inference*, Bd. 65(1) (1997), S. 159–189.
- [35] JAECKEL, L. A.: Some Flexible Estimates of Location. In: *The Annals of Mathematical Statistics*, Bd. 42(5) (1971), S. 1540–1552.
- [36] KOCH, R.: Untersuchung von a-priori Verteilungen für Varianzparameter in einem bayesianischen meta-analytisch prädiktiven Modell zur Extrapolation historischer Ergebnisse für eine neue klinische Studie. In: *GMDs 2013: 58. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e. V. (GMDs). Lübeck, 01. - 05.09.2013*. Hrsg. von HANDELS, H. und INGENERF, J. Düsseldorf: German Medical Science GMS Publishing House, 2013.
- [37] KOCH, R.; GERSS, J.; FALDUM, A. und KÖPCKE, W.: Improvement of a Bayesian model: Combining results from adult and pediatric clinical trials for planning and analyzing a new pediatric trial. In: *58. Biometrisches Kolloquium, March 13–15, 2012, Beuth Hochschule für Technik, Berlin, Programm*. Deutsche Region der Internationalen Biometrischen Gesellschaft (IBS-DR), 2012, S. 162–163.
- [38] KOCH, R.; GERSS, J.; FALDUM, A. und KÖPCKE, W.: Optimierung eines Bayes-Ansatzes zur Nutzung von Ergebnissen aus Erwachsenen- und Kinderstudien für die Planung einer neuen pädiatrischen Studie. In: *Mainz 2011. 56. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (gmds), 6. Jahrestagung der Deutschen Gesellschaft*

- für *Epidemiologie (DGEpi)*. Mainz, 26.-29.09.2011. Düsseldorf: German Medical Science GMS Publishing House, 2011.
- [39] KOCH, R.; GERSS, J.; FALDUM, A. und KÖPCKE, W.: Utilizing historical information from adult and pediatric clinical trials: Results, limitations, and extensions of the meta-analytic predictive approach. In: *GMDS 2012. 57. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e.V. (GMDS)*. Braunschweig, 16.-20.09.2012. Düsseldorf: German Medical Science GMS Publishing House, 2012.
- [40] LAMBERT, P. C.; SUTTON, A. J.; BURTON, P. R.; ABRAMS, K. R. und JONES, D. R.: How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. In: *Statistics in Medicine*, Bd. 24(15) (2005), S. 2401–2428.
- [41] LINK, W. A. und EATON, M. J.: On thinning of chains in MCMC. In: *Methods in Ecology and Evolution*, Bd. 3(1) (2012), S. 112–115.
- [42] LÜLLMANN, H.; MOHR, K. und HEIN, L.: *Pharmakologie und Toxikologie: Arzneimittelwirkungen verstehen - Medikamente gezielt einsetzen*. Thieme Georg Verlag, 2010.
- [43] LUNN, D. J.; THOMAS, A.; BEST, N. und SPIEGELHALTER, D.: WinBUGS – A Bayesian modelling framework: Concepts, structure, and extensibility. In: *Statistics and Computing*, Bd. 10(4) (2000), S. 325–337.
- [44] LUNN, D.; SPIEGELHALTER, D.; THOMAS, A. und BEST, N.: The BUGS project: Evolution, critique and future directions. In: *Statistics in Medicine*, Bd. 28(25) (2009), S. 3049–3067.
- [45] LYNCH, S. M.: *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists (Statistics for Social and Behavioral Sciences)*. 1. Aufl. Springer, 2007.
- [46] MARLEY, J. K. und WAND, M. P.: Non-Standard Semiparametric Regression via BRugs. In: *Journal of Statistical Software*, Bd. 37(5) (2010), S. 1–30.
- [47] MASSICOTTE, P.; JULIAN, J. A.; GENT, M.; SHIELDS, K.; MARZINOTTO, V.; SZECHTMAN, B.; ANDREW, M. und GRP, R. S.: An open-label randomized controlled trial of low molecular weight heparin compared to heparin and coumadin for the treatment of venous thromboembolic events in children: the REVIVE trial. In: *Thromb Res*, Bd. 109(2-3) (2003), S. 85–92.
- [48] METROPOLIS, N.; ROSENBLUTH, A. W.; ROSENBLUTH, M. N.; TELLER, A. H. und TELLER, E.: Equation of State Calculations by Fast Computing Machines. In: *Journal of Chemical Physics*, Bd. 21 (1953), S. 1087–1092.
- [49] MUELLER, U.; NAUCK, B. und DIEKMANN, A.: *Handbuch der Demographie 2: Anwendungen*. Handbuch der Demographie. Springer-Verlag GmbH, 2000.
- [50] NEUENSCHWANDER, B.; CAPKUN-NIGGLI, G.; BRANSON, M. und SPIEGELHALTER, D. J.: Summarizing historical information on controls in clinical trials. In: *Clinical Trials*, Bd. 7(1) (2010), S. 5–18.

- [51] NORMAND, S.-L. T.: Meta-analysis: formulating, evaluating, combining, and reporting. In: *Statistics in Medicine*, Bd. 18(3) (1999), S. 321–359.
- [52] PETSKEY, H. L.; CATES, C. J.; LASSERSON, T. J.; LI, A. M.; TURNER, C.; KYNASTON, J. A. und CHANG, A. B.: A systematic review and meta-analysis: tailoring asthma treatment on eosinophilic markers (exhaled nitric oxide or sputum eosinophils). In: *Thorax*, Bd. 67(3) (2012), S. 199–208.
- [53] R CORE TEAM: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2012.
- [54] RUE, H.; MARTINO, S. und CHOPIN, N.: Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Bd. 71(2) (2009), S. 319–392.
- [55] SACKS, H.; CHALMERS, T. C. und SMITH H., J.: Randomized versus historical controls for clinical trials. In: *American Journal of Medicine*, Bd. 72(2) (1982), S. 233–40.
- [56] SAENZ, A.; FERNANDEZ-ESTEBAN, I.; MATAIX, A.; AUSEJO, M.; ROQUE, M. und MOHER, D.: Metformin monotherapy for type 2 diabetes mellitus. In: *Cochrane Database Syst Rev*, Bd. (3) (2005), S. CD002966.
- [57] SAVCHUK, V. und TSOKOS, C.: *Bayesian Theory and Methods with Applications*. Atlantis Studies in Probability and Statistics. Atlantis Press, 2011.
- [58] SCHOENFELD, D. A.; ZHENG, H. und FINKELSTEIN, D. M.: Bayesian design using adult data to augment pediatric trials. In: *Clinical Trials*, Bd. 6(4) (2009), S. 297–304.
- [59] SMITH, T. C.; SPIEGELHALTER, D. J. und THOMAS, A.: Bayesian approaches to random-effects meta-analysis: A comparative study. In: *Statistics in Medicine*, Bd. 14(24) (1995), S. 2685–2699.
- [60] SPIEGELHALTER, D. J.; THOMAS, A.; BEST, N. G. und GILKS, W. R.: *BUGS Examples Vol.1, Version 0.5 (version ii)*. MRC Biostatistics Unit: Cambridge, 1996.
- [61] SPIEGELHALTER, D. J.; THOMAS, A.; BEST, N. G. und GILKS, W. R.: *BUGS Examples Vol.2, Version 0.5 (version ii)*. MRC Biostatistics Unit: Cambridge, 1996.
- [62] SPIEGELHALTER, D. J.; ABRAMS, K. R. und MYLES, J. P.: *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. John Wiley & Sons, Ltd, 2004.
- [63] STURTZ, S.; LIGGES, U. und GELMAN, A.: R2WinBUGS: A Package for Running WinBUGS from R. In: *Journal of Statistical Software*, Bd. 12(3) (2005), S. 1–16.
- [64] SYVERSVEEN, A.: *Noninformative Bayesian Priors: Interpretation and Problems with Construction and Applications*. Statistics (Trondheim : trykt utg.) Norges teknisk-naturvitenskapelige universitet, 1998.
- [65] TSIATIS, A. A.: The asymptotic joint distribution of the efficient scores test for the proportional hazards model calculated over time. In: *Biometrika*, Bd. 68(1) (1981), S. 311–315.

- [66] U.S. FOOD AND DRUG ADMINISTRATION (FDA): *Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics*. 2010.
- [67] U.S. FOOD AND DRUG ADMINISTRATION (FDA): *Guidance for Industry and FDA Staff: Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials*. 2010.
- [68] VAN DONGEN, C. J.; VAN DEN BELT, A. G.; PRINS, M. H. und LENSING, A. W.: Fixed dose subcutaneous low molecular weight heparins versus adjusted dose unfractionated heparin for venous thromboembolism. In: *Cochrane Database Syst Rev*, Bd. (4) (2004), S. CD001100.
- [69] VAN HOUWELINGEN, H. C.; ARENDS, L. R. und STIJNEN, T.: Advanced methods in meta-analysis: multivariate approach and meta-regression. In: *Statistics in Medicine*, Bd. 21(4) (2002), S. 589–624.
- [70] VIECHTBAUER, W.: Bias and Efficiency of Meta-Analytic Variance Estimators in the Random-Effects Model. In: *Journal of Educational and Behavioral Statistics*, Bd. 30(3) (2005),
- [71] WARN, D. E.; THOMPSON, S. G. und SPIEGELHALTER, D. J.: Bayesian random effects meta-analysis of trials with binary outcomes: methods for the absolute risk difference and relative risk scales. In: *Statistics in Medicine*, Bd. 21(11) (2002), S. 1601–1623.
- [72] WARNES, G. R.: *Various R programming tools*. Includes R source code and/or documentation contributed by Ben Bolker and Thomas Lumley. 2012.

