

Schätzung von Effektgrößen

Empirische Untersuchungen
zu standardisierten Lagedifferenzen und zur stochastischen Überlegenheit
bei unverbundenen und verbundenen Stichproben

Inaugural-Dissertation
zur Erlangung der Doktorwürde
der
Philosophischen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität
zu Bonn

vorgelegt von
Norbert Hilger

Bonn, März 2010

Gedruckt mit Genehmigung der Philosophischen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn

Diese Dissertation ist auf dem Hochschulschriftenserver der ULB Bonn
http://hss.ulb.uni-bonn.de/diss_online elektronisch publiziert.

Erstgutachter: Prof. Dr. Georg Rudinger
Zweitgutachter: Prof. Dr. Arndt Bröder

Tag der mündlichen Prüfung: 26. Februar 2009

Für Geduld, Rat und Freundschaft bedanke ich mich herzlich bei Annette Wille, Prof. Dr. Georg Rudinger, Dr. Michael Rietz (†), Simon Riek, Dr. Thomas Krüger, Svenja Kremer und Dr. Klaus-Martin Klein.

Kurzfassung

Vor dem Hintergrund, daß der Schätzung von Effektgrößen in der psychologischen Forschung ein zunehmender Stellenwert zukommt, werden in der vorliegenden Arbeit für zwei grundlegende Datensituationen, dem Zwei-Gruppen-Fall mit unabhängigen und dem Zwei-Gruppen-Fall mit verbundenen Stichproben, die Eigenschaften von Punkt- und insbesondere Intervallschätzern für standardisierte Lagedifferenzen (u.a. Cohens d) und die stochastische Überlegenheit $\Pr(X_1 \geq X_2)$ anhand von Monte-Carlo-Simulationen untersucht. Das wesentliche Ziel hierbei ist es, geeignete Schätzer für das verallgemeinerte Lagen-Skalen-Modell zu finden, also Verfahren, die sich auch unter nichtnormalen Verteilungen mit heterogenen Varianzen robust verhalten.

In den beiden Studien zu standardisierten Lagedifferenzen werden nichtzentrale sowie Bootstrap-Schätzer für Kleinste-Quadrat-(KQ)- und robustifizierte Effektgrößen (20%-getrimmt/winsorisiert) untersucht. Zentraler Befund ist, daß die herkömmlichen KQ-Parameter unter Verteilungen mit Tailstärken, die oberhalb derjenigen der Normalverteilung liegen, nicht robust geschätzt werden können, sondern das nominale Konfidenzniveau selbst bei großen Stichprobenumfängen unterschreiten. Hingegen gelingt die Schätzung der robustifizierten Größen insbesondere mit dem Perzentil-Bootstrap über alle Verteilungen hinweg relativ robust. Da jedoch die Intervalle der robustifizierten Schätzungen unter Verteilungen mit schwachen Tails deutlich breiter als diejenigen der KQ-Schätzungen ausfallen, wird von einer generellen Auswertung anhand robustifizierter Statistiken abgeraten und eine modellabhängige Wahl des Schätzers empfohlen.

Neben dem Vorteil der generellen Erwartungstreue der Punktschätzer zeigen die Simulationsstudien zur stochastischen Überlegenheit, daß die Intervallschätzer dieser ordinalen Statistik eine vergleichsweise höhere Robustheit hinsichtlich der Einhaltung des Konfidenzniveaus aufweisen als die Verfahren zur Schätzung der Lagedifferenzen. Bei unabhängigen Stichproben kann die stochastische Überlegenheit im stetigen Fall bereits bei sehr kleinen Stichproben mit dem in dieser Arbeit entwickelten, als Shift-Intervall bezeichneten Verfahren in aller Regel robust geschätzt werden. Für diskrete Verteilungen wird das Verfahren nach Mee (1990) empfohlen. Bei verbundenen Stichproben werden in Anlehnung an Cliff (1996a) zwei Parameter unterschieden, wobei sich von den hier untersuchten Verfahren für die Schätzung der intraindividuellen stochastischen Überlegenheit das adjustierte Wald-Intervall nach Agresti und Coull (1998) und für die Schätzung der interindividuellen stochastischen Überlegenheit ein Pseudo-Scoreintervall mit dem Varianzschätzer von Cliff anbieten.

Inhalt

1 Einleitung	1
1.1 Signifikanz und Effekt	1
1.2 Methodologische Aspekte.....	2
1.3 Metaanalyse	5
1.4 Testen und Schätzen	7
1.5 Untersuchungsgegenstand und Aufbau der Arbeit	10
2 Standardisierte Lagedifferenzen	13
2.1 Unverbundene Stichproben.....	13
2.1.1 Parameter und Punktschätzer.....	13
2.1.1.1 Cohens d und d'	13
2.1.1.2 Glass' Δ	16
2.1.1.3 Hedges' erwartungstreuer Schätzer für d	17
2.1.1.4 Huynhs Biaskorrektur bei Varianzheterogenität	17
2.1.1.5 Bootstrap-Biaskorrektur	18
2.1.1.6 Effektgrößen mit getrimmten Mittelwerten	18
2.1.2 Intervallschätzer.....	22
2.1.2.1 Zentrale Schätzer.....	22
2.1.2.2 Nichtzentrale Schätzer	23
2.1.2.3 Bootstrap-Schätzer	24
2.1.2.3.1 Perzentil-Verfahren.....	24
2.1.2.3.2 BCA-Verfahren.....	24
2.1.2.3.3 Perzentil- t -Verfahren	25
2.1.3 Fragestellung und ausgewählte Statistiken.....	26
2.1.3.1 Punktschätzer	28
2.1.3.2 Intervallschätzer	28
2.1.4 Simulationsmethode	30
2.1.4.1 Verwendete Verteilungen	30
2.1.4.2 Erzeugung der Pseudozufallszahlen.....	33
2.1.4.3 Datenmodelle	34
2.1.4.4 Robustheitskriterien und Anzahl der Simulationsdurchgänge.....	35
2.1.4.5 Untersuchungsdesign	37
2.1.5 Ergebnisse der Studie 1	38
2.1.5.1 Punktschätzer	38
2.1.5.1.1 Schätzer für d'	38
2.1.5.1.2 Schätzer für d^{S_1} und d^{S_2}	40
2.1.5.1.3 Schätzer für d'_R	43
2.1.5.1.4 Schätzer für $d_R^{S_1}$ und $d_R^{S_2}$	44
2.1.5.1.5 Vergleich von KQ- und robustifizierter Schätzung	46
2.1.5.2 Intervallschätzer	47
2.1.5.2.1 Schätzer für d'	47
2.1.5.2.2 Schätzer für d^{S_1} und d^{S_2}	52

2.1.5.2.3	Schätzer für d'_R	57
2.1.5.2.4	Schätzer für $d_R^{S_1}$ und $d_R^{S_2}$	61
2.1.5.2.5	Vergleich von KQ- und robustifizierter Schätzung	66
2.1.5.3	Fazit.....	67
2.2	Verbundene Stichproben.....	69
2.2.1	Parameter und Punktschätzer.....	69
2.2.1.1	Cohens d_D	69
2.2.1.2	Cohens d' bei Abhängigkeit	70
2.2.1.3	Glass' Δ	72
2.2.1.4	Effektgrößen mit getrimmten Mittelwerten	72
2.2.2	Intervallschätzer.....	73
2.2.2.1	Exakter Schätzer für d_D	73
2.2.2.2	Nichtzentrale Schätzer für d^{S_k} und d'	73
2.2.3	Fragestellung und ausgewählte Statistiken.....	74
2.2.3.1	Punktschätzer	75
2.2.3.2	Intervallschätzer	75
2.2.4	Simulationsmethode	76
2.2.4.1	Erzeugung der korrelierten Pseudozufallszahlen	77
2.2.4.2	Datenmodelle	77
2.2.4.3	Untersuchungsdesign	77
2.2.5	Ergebnisse der Studie 2	78
2.2.5.1	Punktschätzer	78
2.2.5.1.1	Schätzer für d_D	78
2.2.5.1.2	Schätzer für d'	79
2.2.5.1.3	Schätzer für d^{S_1} und d^{S_2}	80
2.2.5.1.4	Schätzer für $d_{D,R}$	81
2.2.5.1.5	Schätzer für d'_R	82
2.2.5.1.6	Schätzer für $d_R^{S_1}$ und $d_R^{S_2}$	83
2.2.5.2	Intervallschätzer	84
2.2.5.2.1	Schätzer für d_D	84
2.2.5.2.2	Schätzer für d'	87
2.2.5.2.3	Schätzer für d^{S_1} und d^{S_2}	89
2.2.5.2.4	Schätzer für $d_{D,R}$	94
2.2.5.2.5	Schätzer für d'_R	96
2.2.5.2.6	Schätzer für $d_R^{S_1}$ und $d_R^{S_2}$	98
2.2.5.2.7	Vergleich von KQ- und robustifizierter Schätzung	102
2.2.5.3	Fazit.....	104
3	Stochastische Überlegenheit	105
3.1	Unverbundene Stichproben.....	105
3.1.1	Varianten in der psychologischen Literatur.....	105
3.1.1.1	Common Language Effect Size nach McGraw und Wong.....	106
3.1.1.2	Dominance Statistic nach Cliff	107
3.1.1.3	Probability of Superiority nach Grissom.....	107
3.1.1.4	Measure of Stochastic Superiority nach Vargha und Delaney.....	108

3.1.2	Parameter und Punktschätzer.....	108
3.1.3	Intervallschätzer.....	109
3.1.3.1	Methode von Mee	111
3.1.3.2	Cliffs Schätzer für die Dominanzstatistik	112
3.1.3.3	Rang-Welch-Verfahren nach Vargha und Delaney	114
3.1.3.4	Test nach Brunner und Munzel	114
3.1.3.5	Pseudo-Score-Intervall nach Newcombe	116
3.1.3.6	Edgeworth-Reihe der U -Statistik	116
3.1.3.7	Shift-Konfidenzintervall	117
3.1.4	Notation der ausgewählten Statistiken	119
3.1.5	Verteilungsmodelle und Untersuchungsdesign	120
3.1.6	Ergebnisse der Studie 3	121
3.2	Verbundene Stichproben.....	130
3.2.1	Parameter und Punktschätzer.....	130
3.2.2	Intervallschätzer.....	131
3.2.2.1	Stochastische Überlegenheit der intraindividuellen Veränderung.....	131
3.2.2.1.1	Stetige Verteilungen	131
3.2.2.1.1.1	Wald-Intervall.....	131
3.2.2.1.1.2	Score-Intervall nach Wilson	132
3.2.2.1.1.3	Score-Intervall mit Kontinuitätskorrektur nach Blyth und Still ...	132
3.2.2.1.1.4	Adjustiertes Wald-Intervall.....	132
3.2.2.1.1.5	Exaktes Intervall nach Clopper und Pearson	132
3.2.2.1.1.6	Approximation der Binomial-Verteilung nach Pratt	133
3.2.2.1.2	Diskrete Verteilungen	134
3.2.2.1.2.1	Adjustiertes Wald-Intervall.....	135
3.2.2.1.2.2	MAX und MID-P	135
3.2.2.1.2.3	Schätzer nach Cliff.....	135
3.2.2.1.3	Datenbeispiel	136
3.2.2.2	Stochastische Überlegenheit zwischen den Gruppen.....	136
3.2.2.2.1	Schätzer nach Cliff.....	136
3.2.2.2.2	t -Tests auf Basis von Rängen.....	137
3.2.2.2.3	Shift-Konfidenzintervall	138
3.2.3	Notation der ausgewählten Statistiken	139
3.2.4	Untersuchungsdesign.....	140
3.2.5	Ergebnisse der Studie 4	140
3.2.5.1	Schätzer für θ_w	140
3.2.5.2	Schätzer für θ_b	145
4	Anwendungen	151
4.1	Fahreignung und Spurhalten	151
4.2	Progressive Matrizen-Tests mit und ohne Speed-Komponente.....	153
5	Zusammenfassung	157
Literatur	163

Anhang A Symbolverzeichnis und Abkürzungen	173
Anhang B Parameter der Pseudozufallszahlen	175
Anhang C Ergebnisse der Studie 1	179
Anhang D Ergebnisse der Studie 2	211
Anhang E Ergebnisse der Studie 3	259
Anhang F Ergebnisse der Studie 4	271

1 Einleitung

Der Begriff *Effektgröße* wurde von Cohen (1962, 1969) im Zusammenhang mit der statistischen Teststärkeanalyse in die psychologische Literatur eingeführt. Nach Cohen sind Effektgrößen Populationsparameter, die jeweils im Modell eines Tests die Unterschiedlichkeit oder den Zusammenhang von Variablen in standardisierter, metrikfreier Form zum Ausdruck bringen. Prominentes Beispiel einer Effektgröße ist der Index $d = (\mu_1 - \mu_2)/\sigma$, der beim t -Test zweier unabhängiger Mittelwerte zur Anwendung kommen kann. Neben ihrer Funktion in der Teststärkeanalyse dienen Effektgrößen zudem als Schätzstatistiken sowohl im Zusammenhang mit Signifikanztests als auch unabhängig von diesen (Grissom & Kim, 2005) sowie schließlich als (abhängige) Variable in der Metaanalyse (Hunter & Schmidt, 1990; Rosenthal, 1991). Im folgenden wird der theoretische Hintergrund skizziert, vor dem die Entwicklung und Anwendung von als Effektgrößen bezeichneten Maßzahlen stattgefunden hat, sowie die Fragestellung der vorliegenden Arbeit entworfen.

1.1 Signifikanz und Effekt

Da im herkömmlichen Testparadigma mit einer Nullhypothese behauptet wird, daß ein Effekt der Größe Null vorliegt, lassen sich die von Cohen (1988) definierten Effektgrößen als Maß für den Abstand einer einfachen H_1 von der H_0 interpretieren (Buchner, Erdfelder & Faul, 1996), oder auch als Ausmaß bezeichnen, zu welchem die H_0 falsch ist. In der Teststärkeanalyse werden vier Größen behandelt: die Fehlerwahrscheinlichkeiten α und β , der Stichprobenumfang n und die Effektgröße. Diese hängen so zusammen, daß jeweils eine Größe aus den Werten der anderen drei resultiert, wodurch sich zumindest vier Arten von Teststärkeanalysen ergeben (vgl. Faul, Erdfelder, Lang & Buchner, 2007).

Cohens (1969, 1988) Arbeiten zur Teststärke verschiedener parametrischer Tests haben in der Psychologie die Optimierung von Signifikanztests in der Tradition von Neyman und Pearson (1928, 1933) bekannt und anwendbar gemacht. Zur Optimierung eines Signifikanztests muß eine einfache Alternativhypothese oder ein Mindesteffekt formuliert werden, wobei dies vielfach nur in Einheiten einer Effektgröße gelingt. Während die Effektgröße bei manchen Verfahren mit dem zu testenden Parameter identisch oder diesem eindeutig zugeordnet ist, so z.B. beim Gauß-Test, ist sie in der Regel eine Funktion mehrerer Parameter, wobei ein bestimmter Wert einer Effektgröße immer zu einer Klasse von einfachen Alternativhypothesen gehört, aus denen sich jeweils dieselbe Verteilung der Teststatistik ableiten läßt. Je mehr Parameter also in einer Effektgröße verrechnet werden, desto weniger sagt diese über die Skala der abhängigen Variablen aus. Ein großer Wert einer Effektgröße wie z.B. d kann gleichermaßen auf einen großen Effekt (systematische Varianz) wie auch auf eine geringe Binnenvarianz zurückgehen. Im Rahmen des parametrischen Testmodells stehen gleiche Werte einer Effektgröße für gleichartige Populationsgegebenheiten, und je größer der Wert oder Betrag einer Effektgröße ist, desto größer sind die Unterschiede oder Zusammenhänge der Variablen.

Zwar beschäftigt sich Cohen nur wenig mit den statistischen Aspekten der Schätzung von Effektgrößen, doch er sieht ihre Funktion keineswegs beschränkt auf den Einsatz als standardisierte Alternativhypothese. Vielmehr empfiehlt er die Schätzung von Effektgrößen sowohl in Ergänzung zu als auch unabhängig von der Durchführung von Signifikanztests (1965, S. 101 f.; 1988, S. 68). Während die Verwendung von Effektgrößen als eigenständige Schätzstatistiken erst seit der zweiten großen Signifikanztestdebatte in den neunziger Jahren weitere Verbreitung findet (Harlow, Mulaik & Steiger, 1997), gehört die Schätzung einer Effektgröße in Ergänzung zum Signifikanztest zum

Methodenstandard in der Psychologie. Wesentlicher Grund hierfür ist, daß in der Regel eine Optimierung des Signifikanztests im Sinne von Neyman und Pearson nicht durchgeführt wird, sondern eine Testung nach Fisher (1925, 1935) mit zusammengesetzter Alternativhypothese erfolgt. Da in der Psychologie aber die Ablehnung einer Nullhypothese nicht unbedingt als wissenschaftlich bedeutsame Entscheidung beurteilt wird (Bredenkamp, 1972), wird allgemein gefordert, neben der Testentscheidung das Ausmaß des gegebenen Unterschieds oder Zusammenhangs mitzuteilen. Anders als von Fisher vorgesehen, hat sich hierfür der p -Wert nicht durchgesetzt, da dieser von der Stichprobengröße abhängt. In Lehrbüchern findet sich vielfach folgende Argumentation: Da weder die Annahme der Alternativhypothese für einen bedeutsamen Effekt noch die Beibehaltung der Nullhypothese für einen irrelevanten Effekt sprechen muß, soll zusätzlich die Größe des gefundenen Effekts geschätzt werden. Die hierfür empfohlenen Statistiken wurden auch als Maße der praktischen Bedeutsamkeit bezeichnet. In dieser Bezeichnung kommt die Problematik des Signifikanztests zwar gut zum Ausdruck, sie ist aber etwas irreführend, da die Bedeutsamkeit auch mit diesen Maßen allein noch nicht erfaßt wird; hierfür ist zusätzlich eine Referenz zur Beurteilung des numerischen Wertes erforderlich. Die Bedeutsamkeit eines Ergebnisses kann letztlich nur inhaltlich, im jeweiligen fachwissenschaftlichen Kontext vorgenommen werden. Im folgenden wird daher neutraler von Effektgrößen und Schätzern von Effektgrößen gesprochen. Das skizzierte Vorgehen stellt bereits eine Anreicherung des Signifikanztests nach Fisher um ein Konzept der Neyman-Pearson-Theorie dar, denn nach Fisher kann einzig der p -Wert über die Bedeutsamkeit eines Unterschiedes Auskunft geben (Fisher, 1925; Hager, 1992). Die in der psychologischen Forschung vielfach vorgenommene Vermischung der ursprünglich konkurrierenden statistischen Testtheorien hat nach Gigerenzer (1993) zu einer fehlerhaften, dem wissenschaftlichen Fortschritt gar entgegenstehenden Hybridlogik geführt; daß aber eine Verbindung beider Ansätze prinzipiell möglich ist und eventuell ein überlegenes Verfahren erbringt, hat Lehmann (1993) dargestellt.

Der Zusammenhang von Effektgröße und Signifikanztest – mit *dem* Signifikanztest sind im folgenden die Verfahren sämtlicher Testtheorien gemeint, sofern eine konzeptuelle Unterscheidung nicht erforderlich erscheint – kommt besonders in der Definition von Tatsuoka (1993) zum Ausdruck:

$$\text{Effektgröße} = \text{Signifikanzteststatistik} / \text{Stichprobenumfang} .$$

Da die Effektgröße oben als Parameter eingeführt wurde, müßte auf der linken Seite richtigerweise Schätzer der Effektgröße stehen. Für den Schätzer für d besteht z.B. folgende Beziehung: $\hat{d} = t / \sqrt{n_1 n_2 / (n_1 + n_2)}$, wobei n_1 und n_2 die beiden Gruppengrößen bezeichnen.

1.2 Methodologische Aspekte

Im deutschen Sprachraum haben sich vor allem Bredenkamp (1969, 1972, 1980) sowie mehrere seiner Schüler (Erdfelder, 2004; Hager, 1992, 2004; Westermann, 2000) mit der methodologischen Bedeutung des Hypothesentests nach Neyman und Pearson und in diesem Zusammenhang mit der Funktion von Effektgrößen beschäftigt. Hinsichtlich der konkreten Gestaltung einer Teststärkeanalyse weichen diese Autoren nicht wesentlich von Cohens (1988) Darstellung ab, jedoch folgt aus der experimentellen Praxis eine andere Interpretation des parametrischen Tests. Da in der Psychologie in aller Regel keine Zufallsstichproben gezogen werden (vgl. Hager, 2005) und zudem die Verteilungsannahmen der parametrischen Tests fragwürdig erscheinen, greifen sie Edgingtons (1966) Interpretation der parametrischen Tests als approximative Randomisationstests auf. Denn bei

einem Randomisationstest muß nicht davon ausgegangen werden, daß die Untersuchungseinheiten aus einer tatsächlichen oder hypothetischen Population zufällig (und mit Zurücklegen) gezogen wurden. Vielmehr erfordert der Randomisationstest lediglich die Unabhängigkeit der Einheiten sowie ihre zufällige Aufteilung auf die Untersuchungsbedingungen. Diese weniger restriktiven Voraussetzungen sind in der experimentellen Arbeit vergleichsweise einfach sicherzustellen. Die genannte Interpretation des parametrischen Tests hat aber auch Folgen für die Bedeutung der empirischen Effektgröße. Da es bei einer nicht zufälligen Stichprobe nur um den konkret beobachteten Effekt geht und eben nicht auf eine wie auch immer geartete Population geschlossen werden soll, fassen die Autoren auch die beobachtete Effektgröße nicht als Schätzwert auf: „Wir favorisieren \hat{R}^2 bzw. das korrigierte \hat{R}^2 , weil dieses Maß nicht vortäuscht, eine *Schätzung* für einen Parameter zu sein. Es genügt zu wissen, wie groß der Zusammenhang für ein bestimmtes Experiment ist, da im Rahmen der deduktivistischen Theorie keine Verallgemeinerungen auf Populationen angestrebt sind“ (Bredenkamp, 1980, S. 52 f., Hervorhebung im Original). In ähnlicher Weise äußert sich auch Hager (1996, S. 93), wobei hier deutlicher wird, daß die im folgenden Zitat genannten Autoren schätztheoretische Konzepte wie etwa Erwartungstreue keineswegs unbeachtet lassen: „Wenn man an der Schätzung des $\eta_{Y,A}^2$ in der zugrundeliegenden Population interessiert ist, dann erweist es sich, daß das berechnete multiple Korrelationsquadrat das ‚wahre‘ multiple Korrelationsquadrat in der Population überschätzt. Zum Ausgleich dieser Überschätzung sind verschiedene sog. ‚Schrumpfungskorrekturen‘ vorgeschlagen worden, deren wichtigste sich u.a. bei Bredenkamp [...], bei Hager und Westermann [...] und bei Thompson [...] finden. Aber diese Schätzung von Populationsmaßen ist bei der Prüfung von psychologischen Hypothesen über aus ihnen abgeleitete statistische Hypothesen [...] weniger bedeutungsvoll als in der herkömmlichen Inferenzstatistik [...]“. Obwohl die genannten Autoren nicht auf die Schätzung von Parametern abzielen, schlagen sie aber offensichtlich verzerrungskorrigierte Statistiken vor. Ein Grund hierfür mag darin liegen, daß es keinen anderen statistischen Bezugsrahmen als die herkömmliche Schätztheorie gibt, der die Unabhängigkeit einer Statistik von der Stichprobengröße beurteilen ließe. Und die Unabhängigkeit der Effektgröße von n , sowohl im Hinblick auf die Gesamtstichprobengröße als auch auf die Verhältnisse der Gruppengrößen, ist für diese Autoren wesentlich. Denn anhand empirischer Befunde sollen zum einen Effektgrößen zur Teststärkeanalyse für nachfolgende Untersuchungen festgelegt werden, wo sie dann als *Parameter* der Alternativhypothese fungieren, und zum anderen soll der Vergleich zwischen Untersuchungen Aufschluß über die Durchschlagskraft unabhängiger Variablen geben. Für die Hypothesenprüfung selbst hat die angesprochene Schätzproblematik, nachdem eine Effektgröße (a priori) zur Testplanung gewählt wurde, tatsächlich keine Bewandnis mehr. In dieser in die deduktivistische Methodologie eingebetteten Tradition der Hypothesenprüfung findet sich also der auf den ersten Blick verwirrende Umstand, daß man wegen Ermangelung zufälliger Stichproben nicht mittels klassischer Inferenz auf Grundgesamtheiten rückschließen kann, dennoch aber auf deren Verfahren zurückgreift. Auch wenn es also in Untersuchungen ohne zufällige Stichproben nicht um die Schätzung von Parametern definierter Grundgesamtheiten gehen kann, sondern vielmehr um sogenannte Schlüsse auf experimentelle Effekte, so sind aber die schätztheoretischen Eigenschaften der Statistiken wie z.B. Erwartungstreue dennoch wesentlich, wenn Vergleiche zwischen Untersuchungen mit unterschiedlichen Stichprobengrößen vorgenommen oder in der Teststärkeanalyse Effektgrößen anhand verwandter Studien festgelegt werden sollen.

Zu sehr deutlichen Vorbehalten hinsichtlich der quantitativen Information eines beobachteten Effekts kommt Chow (1988). Zunächst unterscheidet Chow zwei Arten der experimentellen Forschung: die deskriptive und die theoriestende. Bei einem deskriptiven Experiment liegt das

wissenschaftliche Interesse unmittelbar in der Wirksamkeit der experimentellen Manipulation. In diesem als agrarwissenschaftlich bezeichneten Wissenschaftsmodell gibt es keine Unterscheidung zwischen theoretischer und praktischer Fragestellung. Ob ein Dünger wirkt oder nicht, wird getestet, indem eben dieser Dünger eingesetzt und der Ernteerfolg gemessen wird. Während in der deskriptiven Forschung die Frage nach der Größe eines Effekts zentral ist, gibt es hingegen in der theoriebestimmenden Forschung kein unmittelbares Interesse an der empirischen Fragestellung und den resultierenden Daten. Denn die experimentelle Manipulation ist nur eine von vielen möglichen Operationalisierungen, die wiederum nur eine von vielen Implikationen der nicht beobachtbaren theoretischen Aussagen (zudem unter der erforderlichen Annahme weiterer Hilfhypothesen) betrifft. Die quantitative Größe eines Effekts hängt nicht notwendig mit der Bedeutung einer wissenschaftlichen Hypothese zusammen. Die Aufgabe der Statistik liegt hier allein in der Herbeiführung der binären Entscheidung, ob die Variation in den Daten zufällig oder systematisch ist. Die Erfordernis einer binären Entscheidung wird mit der zugrundeliegenden syllogistischen Argumentationsstruktur begründet. In der Größe des beobachteten Effekts liegt nach Chow aber im Vergleich zum Signifikanztest keine zusätzliche Information für die vorläufige Entscheidung über Falschheit oder Richtigkeit einer Hypothese und der dahinterstehenden Theorie. „Unlike subscribers to the agricultural model, the investigator is not interested in the experimental question for its own sake. [...] As has been shown, all that is required of a statistical analysis is a binary decision. This is the case because the validity of the syllogistic argument requires only that information. [...] The paradigm ensures that the theoretical conclusion follows logically from its premises, and logical validity is an all-or-none property of an argument. [...] Consequently, the magnitude of the effect size in an experiment is not necessarily a quantitative index of a theoretical property of the underlying mechanism. An investigator may be misled, however, to think otherwise if the emphasis is on the magnitude of the effect-size estimate“ (Chow, 1988, S. 108 f.).

Im Gegensatz zu Chow und auch Erdfelder, der die Bedeutung der Stichprobeneffektstärke als fraglich beurteilt (2004, S. 119), weist Hager (2004) der beobachteten Effektgröße eine wesentliche Funktion in der Rückschlußkette für die Entscheidung über die psychologische Vorhersage und die psychologische Hypothese zu. Nach der Untersuchung soll zunächst die Entscheidung über die statistischen Hypothesen wie üblich anhand des Vergleichs zwischen kritischem und empirischem Wert getroffen werden. Die Urteile über das Eintreten der vorgeordneten psychologischen Vorhersage sowie über die Bewährung der psychologischen Hypothese sind aber nicht allein abhängig von dieser binären Testentscheidung, sondern werden auch unter Berücksichtigung des Vergleichs zwischen kritischer und beobachteter Effektgröße getroffen. Hierbei sind für die verschiedenen Ergebnismuster, die aus Signifikanz und Größe des Effekts resultieren, unterschiedliche Bewährungsgrade der psychologischen Hypothese möglich, nämlich die Bewährung, die bedingte Bewährung und die Nicht-Bewährung. „Der Effektgröße kommt daher ersichtlich eine besondere Bedeutung zu, denn sie weist von allen Determinanten des statistischen Tests noch den engsten Bezug zu den psychologischen Inhalten auf. Auf Grund dieses Sachverhaltes fungiert die Effektgröße als Hauptkriterium bei der Entscheidung über die psychologische Vorhersage (Effektgröße als Entscheidungskriterium)“ (S. 76).

1.3 Metaanalyse

Werden Effektgrößen im Zusammenhang des Signifikanztests in der Regel auf Ebene der individuellen Studie diskutiert, so hat Glass (1976) die Bedeutung von Effektgrößen für die statistische Integration mehrerer Studien herausgestellt. Diese von Glass Metaanalyse genannte Methodik bezeichnet die quantitative Analyse von quantitativen Analysen, also die weitere Verrechnung von statistischen Ergebnissen mehrerer Einzelstudien. Obwohl Verfahren zur statistischen Integration von Primärstudien bereits sehr viel früher entwickelt wurden (Cochran, 1937; Tippet, 1931), hatten diese nur selten Eingang in die psychologische Literatur gefunden. Der Anspruch der Metaanalyse liegt nach Glass nun darin, die scheinbare Widersprüchlichkeit von Befunden aufzuklären und das in der wissenschaftlichen Informationsflut versunkene Wissen zu bergen. Dieses Unterfangen scheint ihm weder mit den bis heute gebräuchlichen, eher narrativen Überblicksarbeiten noch mit dem Auszählen von Signifikanzen (vote counting) möglich. Bedarf und Relevanz von Metaanalysen werden damit begründet, daß zum einen psychologische Effekte insbesondere außerhalb des Labors fragil sind und eine Forschungsfrage nicht mit wenigen Untersuchungen zu beantworten ist und daß zum anderen bereits zu vielen Themen hinreichend viele Untersuchungen durchgeführt wurden, weshalb die Aufarbeitung der vorhandenen Befunde dringlicher ist als das Hinzufügen weiterer Primärstudien. Da die Sammlung der Primärdaten aus fremden Studien vielfach nicht möglich ist, muß die Metaanalyse bei den in der Literatur verfügbaren Statistiken ansetzen. Hierfür schlägt Glass die Verwendung von Effektgrößen vor, da diese nicht nur für die meisten Arbeiten auch nachträglich bestimmt werden können, sondern vor allem auch das Problem lösen, daß in der Psychologie keine einheitlichen Meßinstrumente verwendet werden. Über die Mittelung von Effektgrößen vergleichbarer Studien sollen außer Existenz und Richtung auch die Größe von Treatmenteffekten ermittelt werden sowie über gruppierte Mittelwerte die Studienfaktoren untersucht werden, von denen die Größe des Effekts abhängt.

Anders als im Rahmen des Hypothesentests werden Effektgrößen bei der Metaanalyse also aus einer schätztheoretischen Perspektive betrachtet. Zudem spielen auf dieser Ebene die Teststärken und Entscheidungen der einzelnen Signifikanztests kaum mehr eine Rolle. Von manch rigorosem Metaanalytiker wird die Bedeutung einer Primärstudie nur noch hinsichtlich ihrer Verwertbarkeit in künftigen Metaanalysen gesehen: „Because of the effects of artefacts such as sampling error and measurement error, it would be more accurate to say that data come to us encrypted, and to understand their meaning we must first break the code. Doing this requires meta-analysis. Therefore any individual study must be considered only a single data point to be contributed to a future meta-analysis” (Schmidt, 1992, S. 1179).

Metaanalyse umschreibt weniger eine einheitliche Methodik als vielmehr ein sich entwickelndes Forschungs- und Anwendungsgebiet mehrerer fachwissenschaftlicher Disziplinen. In der Psychologie lassen sich zumindest drei Arten von Metaanalysen unterscheiden, das sind erstens Analysen, die im wesentlichen deskriptive Methoden verwenden, also die Verteilung der berichteten Effektgrößen darstellen (Glass, McGaw & Smith, 1981), zweitens metaanalytische Methoden, die den Stichprobenfehler der jeweiligen Studie berücksichtigen (Hedges & Olkin, 1985) und schließlich drittens psychometrische Metaanalysen, bei denen ferner Faktoren wie z.B. Reliabilität und Validität der Instrumente berücksichtigt werden, welche die Größe der Effekte beeinflussen (Hunter & Schmidt, 1990). Da diese grobe Dreiteilung eher verschiedene Forschungsrichtungen und Autorengruppen unterscheiden läßt und weniger zur Charakterisierung einer durchgeführten Metaanalyse dient, in welcher eine Auswahl diverser Methoden eingesetzt wird, haben Beelmann und Bliesener (1994) trennschärfere Differenzierungsmerkmale vorgeschlagen. Gemeinsam aber ist all diesen

Varianten die Wahl des standardisierten Effekts der einzelnen Studie als abhängige Variable, wobei insbesondere d oder r Anwendung finden. Als unabhängige oder moderierende Variablen gehen je nach Forschungsbereich, metaanalytischer Fragestellung und Homogenität der Effekte diverse Merkmale der Probanden, des Treatments, des Designs und der Meßinstrumente wie z.B. Alter der Patienten, Schweregrad der Erkrankung, Dauer der Behandlung und Untersuchungszeitraum in die Analyse ein und werden auf ihren Beitrag zur Aufklärung der Varianz der Studieneffekte hin untersucht. Metaanalytische Methoden lassen sich mittlerweile nicht nur als wissenschaftlich etabliert, sondern vor allem auch als zunehmend einflußreich für administrative Entscheidungen im klinisch-psychologischen und medizinischen Feld bezeichnen (Hoffert, 1997). Während die Notwendigkeit von integrativen Studien zur Aufarbeitung eines Forschungsbereiches ganz außer Frage steht, ist die Leistungsfähigkeit von Metaanalysen keineswegs unumstritten. Denn eine Vielzahl von statistischen und methodischen Voraussetzungen müssen erfüllt sein, so daß Effektgrößen verschiedener Studien überhaupt vergleichbar sind und ihre Mittelwerte sinnvoll interpretiert werden können.

Bei der statistischen Aggregation und Analyse der Homogenität der Effektgrößen kommen zwei Modellklassen zum Einsatz, Modelle mit festen oder mit zufälligen Effekten (Hedges & Olkin, 1985). Bei der ersten Variante wird davon ausgegangen, daß die beobachteten Effektgrößen aller Studien Schätzungen desselben festen Parameters sind. Bei der zweiten Variante wird, vergleichbar mit dem varianzanalytischen Modell zufälliger Faktoren oder auch mit der klassischen Testtheorie, von einem zweistufigen Prozeß ausgegangen. Danach liegt jeder Studie ein anderer, zufällig gezogener Effektparameter zugrunde, aus dessen jeweiliger Verteilung wiederum zufällig die zu beobachtende Effektgröße resultiert. Im Vergleich zum Modell mit festen Effekten kommt mit der Varianz zwischen den Studieneffekten also eine weitere Variationsquelle hinzu. Im Gegensatz zur Varianzanalyse werden aber in beiden Modellen die Stichprobenfehler der Effektgrößen, also die Varianzen innerhalb der Studien, nicht als gleich angenommen, weshalb die Entwicklung robuster Schätz- und Teststatistiken hier größere Schwierigkeiten bereitet. So halten die derzeit gängigen Homogenitätstests bei geringer Studienanzahl und kleinen bis mittleren Stichproben der Primärstudien das nominale α -Niveau auch dann nicht ein, wenn die Modellannahmen erfüllt sind (Hartung & Knapp, 2003). Neben der Unabhängigkeit der Effektgrößen wird vor allem von Normalverteilung ausgegangen, sowohl für die Effektgröße jeder Primärstudie als auch für die (Super-)Verteilung der Effektparameter im Modell zufälliger Effekte.

Vielfach vorgebrachte Einwände gegen Metaanalysen sind das „Müll rein, Müll raus“- , das „Abhängigkeits“- und das „Äpfel und Birnen-Problem“ (Fricke & Treinies, 1985). Das erste Problem ergibt sich aus der unterschiedlichen Qualität der berücksichtigten Studien. Methodisch schwächere Arbeiten, denen vor allem eine geringere interne Validität zugesprochen werden muß, sollten mit geringerem Gewicht in die Mittelung eingehen als methodisch saubere Untersuchungen. Das zweite Problem entsteht, wenn aus derselben Stichprobe einer Untersuchung mehrere Effektgrößenschätzungen vorliegen; diese enthalten eventuell jeweils relevante Information für die Fragestellung, dürfen aber nicht einzeln derselben Verrechnung zugeführt werden. Das dritte Problem schließlich bezieht sich auf die Einheitlichkeit der unabhängigen und insbesondere der abhängigen Variablen der betrachteten Untersuchungen. Während die Vielfalt der Treatmentgestaltungen zwar die Genauigkeit metaanalytischer Aussagen einschränkt und die Bildung diverser Subgruppen erforderlich macht, so spricht diese nicht grundsätzlich gegen den metaanalytischen Ansatz. Die Frage aber, ob die abhängigen Variablen verschiedener Studien dasselbe Konstrukt erfassen, ist eine zentrale für die Rechtfertigung einer Metaanalyse. Natürlich ist es vollkommen sinnlos, über (in-

haltlich) verschiedene Variablen zu aggregieren. Die Konstruktvalidität der vielfach nicht standardisierten Meßinstrumente ist im nachhinein jedoch kaum überprüfbar.

Bezüglich der Vergleichbarkeit von Effektgrößen ist danach zu fragen, von welchen manipulierbaren, kontrollierbaren und nicht kontrollierbaren Faktoren einer Untersuchung die Größe eines Treatmenteffekts abhängt. Nach Hunter und Schmidt (1990) sind vor allem mangelnde Validität und Reliabilität der Meßinstrumente, künstliche Dichotomisierung, Variationsbeschränkung sowie Konfundierung durch Drittvariablen verantwortlich für die Verzerrung des zu beobachtenden Effekts. Während diese Störeffekte zu einem gewissen Grad im nachhinein korrigierbar sind, zählt z.B. Hager (2000) für den Bereich der psychologischen Interventionsforschung folgende Einflußfaktoren der Effektgröße auf, die sich einer psychometrischen Bewertung entziehen: Spezifität eines Interventionsprogramms, inhaltliche Nähe des Kriteriums zum Programm, Ähnlichkeit von Interventions- und Erhebungssituation, zeitlicher Abstand zum Interventionsende, Treatment der Vergleichsgruppe, Ausgangsstatus der Personen, Erfahrung der Intervenierenden, Größe der Gruppe und institutionelle Rahmenbedingungen. Aufgrund dieser Vielzahl relevanter Randbedingungen kommt Hager zu dem Urteil, „dass die Werte von Effektgrößen über mehrere Untersuchungen, ja sogar in einer Untersuchung mit mehreren Kriteriumsmaßen oder abhängigen Variablen trotz einer möglicherweise ‚gemeinsamen Metrik‘ nicht miteinander vergleichbar sind“ (S. 161). Weitere Komplikationen ergeben sich aus der Wahl des Untersuchungsdesigns, der Balance der Gruppengrößen und der Heterogenität der Varianzen. Die genannten Probleme lassen sich analog auch in der experimentellen Laborforschung aufzeigen, in der zudem die Homogenität der Stichproben nur selten kontrolliert wird. Während die aufgelisteten Einflußfaktoren der Effektgröße in einer Integrationsstudie mit einer Moderatorenanalyse untersucht werden können, fordern sie für den direkten Vergleich zweier oder mehrerer Untersuchungen anhand der Effektgröße zu großer Vorsicht auf, insbesondere wenn Aussagen über die Effizienz von Treatments angestrebt werden. Es mag trivial klingen, aber der Vergleich zweier Untersuchungen kann nie allein statistisch vorgenommen werden.

1.4 Testen und Schätzen

Die nach wie vor herausragende Stellung des Signifikanztests in der psychologischen Forschung und die Definition der Effektgröße innerhalb des Signifikanztests lassen leicht übersehen, daß die Verwendung von standardisierten Zusammenhangs- und Unterschiedsmaßen bereits eine lange Tradition hat und ihre Entwicklung parallel zu der des Signifikanztests verlief. Ein Beispiel hierfür ist die Korrelation, die in der Mitte des 19. Jahrhunderts entwickelt wurde und in ihrer heutigen Form von Pearson 1905 definiert wurde (Stigler, 1986). Doch auch Unterschiedsmaße wie d und hieraus abgeleitete Überlappungsmaße sind schon lange bekannt (Kelley, 1920; Tilton, 1937). Anders als von Cohen angeregt und wie vor dem Siegeszug des Signifikanztests mitunter üblich (Gigerenzer, 1993; Huberty, 2002), wurden Effektgrößen, mit Ausnahme der bivariaten Zusammenhangsmaße, selten als eigenständige Schätzstatistiken eingesetzt, sondern zumeist nur im Anschluß eines (signifikanten) Testergebnisses berichtet. Effektgrößen wie z.B. Klassifikations- und Überlappungsmaße (Huberty & Lowman, 2000; Wilcox & Muska, 1999), die nur mittelbar mit Signifikanzwerten verknüpft sind, werden praktisch nicht eingesetzt.

Aus den diversen Debatten um den Signifikanztest heraus (Harlow, Mulaik & Steiger, 1997; Morrison & Henkel, 1970) sowie aufgrund einer von der APA eingerichteten Arbeitsgruppe zur Inferenzstatistik wurde der grundsätzlichen Empfehlung zum Einsatz von Effektgrößen erneut Nachdruck verliehen. Das Neue an dieser alten Empfehlung ist, daß die Schätzung der Effektgröße

grundsätzlich und auch unabhängig von der Durchführung eines Signifikanztests angeraten und ihre besondere Bedeutung für die Forschungsintegration herausgestellt wird. Zudem soll zu einer Punktschätzung möglichst immer auch eine Intervallschätzung erfolgen: „Always present effect sizes for primary outcomes. [...] Interval estimates should be given for any effect sizes involving principal outcomes. [...] Comparing confidence intervals from a current study to intervals from previous, related studies helps focus attention on stability across studies [...]. Collecting intervals across studies also helps in constructing plausible regions for population parameters” (Wilkinson & the APA Task Force on Statistical Inference, 1999, S. 599). Die Entscheidung zwischen standardisierten und nichtstandardisierten Maßen soll aufgrund der praktischen Bedeutung der Maßeinheit gefällt werden: „If the units of measurement are meaningful on a practical level (e.g., number of cigarettes smoked per day), then we usually prefer an unstandardized measure (regression coefficient or mean difference) to a standardized measure (r or d)” (S. 599). Da die Verwendung absolut skaliert oder geeichter Variablen eher die Ausnahme ist, kommen in aller Regel standardisierte Maße in Betracht. Diese Empfehlungen haben mittlerweile Einzug gehalten in die Publikationsrichtlinien einer Vielzahl amerikanischer Zeitschriften insbesondere im Bereich der angewandten Sozialwissenschaften (Capraro & Capraro, 2002). Ähnliche Kriterien hat es zwar in der Zeitschrift für Sozialpsychologie bereits 1970 gegeben (Bredenkamp & Feger, 1970), jedoch wurde diesen kaum je Genüge geleistet (Hager & Westermann, 1983).

Das Konfidenzintervall einer Effektgröße wird von manchen Autoren als überlegene Alternative zum Signifikanztest diskutiert (Steiger & Fouladi, 1997). Auch Cohen hat in seinen späteren Arbeiten nicht nur verstärkt auf die Schwächen des Signifikanztests hingewiesen und für einen vermehrten Einsatz von Konfidenzintervallen plädiert (1990, 1994), es werden ihm sogar Äußerungen des Bedauerns nachgesagt, sich zu lange mit Teststärke beschäftigt und nicht früher über Konfidenzintervalle geforscht zu haben (Cumming & Finch, 2001, S. 564). Konfidenzintervallen von Effektgrößen werden dem Signifikanztest gegenüber folgende Vorteile zugesprochen: (1) Konfidenzintervalle sind einfach zu verstehen und erleichtern die Dateninterpretation, (2) sie enthalten sämtliche Informationen der zugehörigen Tests, (3) zusätzlich geben sie die Präzision der Untersuchung wieder und schließlich (4) erleichtern sie den Vergleich von Untersuchungen.

Der erste Punkt richtet sich vor allem an die Vielzahl von Fehlschlüssen, zu denen der Signifikanztest einlädt und wie sie selbst von Wissenschaftlern immer wieder gezogen werden (Haller & Krauss, 2002; Oakes, 1986). Zu entgegnen ist aber, daß zum einen auch das Konfidenzintervall fehlinterpretiert werden kann und daß zum anderen die Interpretation von Daten mit Konfidenzintervallen mitunter schwieriger ist als mit Tests. Ein Grund für die Beliebtheit von Tests ist ja gerade die Klarheit der Entscheidung und Einfachheit der Information. Daß z.B. ein Unterschied signifikant ist, ist einfacher zu verarbeiten als die Information, daß der Unterschied 0.2 beträgt und das 95%-Intervall von 0.05 bis 0.35 reicht.

Die Aussage des zweiten Punktes gilt nur für die Beziehung zwischen einem Intervall und dem Test, bei welchem mit der Nullhypothese ein ebensolcher Effekt von Null behauptet wird. In diesem Fall überdeckt ein Konfidenzintervall der Effektgröße genau dann den Wert Null, wenn der Test für die Beibehaltung der Nullhypothese entscheidet. Im Vergleich zu einem Test nach Neyman und Pearson liegt der wesentliche Unterschied darin, daß das Konzept der Teststärke nicht auf ein Konfidenzintervall übertragbar ist, jedenfalls sofern man die Intervallschätzung als Alternative zum Testen und nicht als ergänzende Methode betrachtet. Für die Planung einer Untersuchung kann zwar unter bestimmten Vorannahmen eine gewünschte Intervallbreite die Wahl der Stichprobengröße steuern, jedoch hat die gewünschte Präzision einer Schätzung wenig zu tun mit dem vorab

behaupteten Effekt. Während bei einem Test zuvor angenommen wird, daß ein Effekt eine bestimmte Größe hat, wird mit einem Intervall nachher untersucht, wie groß der Effekt sein mag. Insofern wird hier also die Einschätzung geteilt, daß ein Intervall nicht und ganz besonders nicht im nachhinein als Test von einfachen statistischen Hypothesen benutzt werden sollte. Das Argument, daß theorietestende Forschung des Signifikanztests bedarf und allein schätztheoretische Ansätze bestenfalls etwas für die explorative Forschung sind, greift aber daneben. Denn wie auch bei den obigen Ausführungen von Hager deutlich geworden sein sollte, ist ein Experimentator zwar bereit, eine statistische (Forschungs-)Hypothese zu verwerfen, aber die zugeordnete wissenschaftliche Hypothese wird deshalb noch nicht als nicht bewährt beurteilt, sondern z.B. anhand der beobachteten Effektgröße weiter untersucht. Warum dann also testen, wenn der Testentscheidung nur bedingt – oder gar nicht, wenn nämlich an der Richtigkeit gezweifelt wird – Folge geleistet wird? Da Gott die 6% genau so liebt wie die 5% (Rosnow & Rosenthal, 1989) und der Psychologe üblicherweise nicht in der Qualitätskontrolle täglich dasselbe Experiment durchführt, um echte, folgenreiche Entscheidungen zu treffen, erscheint die grundsätzliche Frage nach der Geeignetheit statistischer Hypothesentests für psychologische Forschung durchaus angebracht.

Die Punkte drei und vier weisen auf die wesentlichen Vorteile der Intervallschätzung im Vergleich zur Information eines Testergebnisses hin, bei dem die Varianz der Effektgröße in der Regel nicht mitgeteilt wird. Die prinzipielle Vergleichbarkeit von Untersuchungsergebnissen vorausgesetzt, bietet die Betrachtung von Effektgrößenintervallen Vorteile gegenüber dem Vergleich von Testentscheidung und Punktschätzung der Effektgröße, denn in der Breite des Intervalls kommt die Präzision der Untersuchung ebenfalls in standardisierter Form zum Ausdruck.

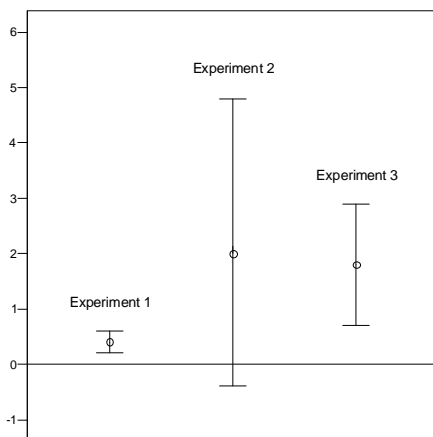


Abbildung 1. Konfidenzintervalle von Effektgrößen (nach Steiger & Fouladi, 1997)

In Abbildung 1 sind die Konfidenzintervalle der Effektgrößen von drei unabhängigen Experimenten wiedergegeben. In den Experimenten 1 und 3 wird die Null jeweils nicht von den Intervallen eingeschlossen, also würden die äquivalenten zweiseitigen Tests für die Annahme der H_1 entscheiden. In Experiment 2 wird die Null hingegen vom Intervall umschlossen, die H_0 wäre hier beizubehalten. Ein Vergleich der Testentscheidungen würde also ergeben, daß die Experimente 1 und 3 zu demselben Ergebnis gekommen sind, während in Experiment 2 kein Effekt gezeigt werden konnte. Die Konfidenzintervalle legen eine andere Interpretation nahe. Im ersten Experiment wurde mit hoher Präzision ein kleiner Effekt entdeckt. Mit Präzision ist hier die relative Breite des Intervalls

gemeint, die auf das Zusammenspiel von Varianz und Stichprobengröße zurückgeht. In Experiment 2 wurde zwar der größte Effekt gefunden, doch aufgrund zu weniger Versuchspersonen, einer heterogenen Stichprobe oder auch eines weniger reliablen Instrumentes, wird dieser Effekt nur sehr ungenau geschätzt. Experiment 3 zeigt einen eher großen Effekt bei einer mittleren Schätzgenauigkeit. Insgesamt deuten die Intervalle darauf hin, daß alle untersuchten Effekte (eher) positiv sind, daß Experiment 1 und Experiment 3 auf verschiedenen große Effekte hindeuten und also nicht dieselben Ergebnisse erbracht haben, und schließlich daß Experiment 2 zu ungenau ist, als daß eine vorläufige Zusammenfassung mit den anderen Ergebnissen möglich wäre. Zu beachten ist hierbei, daß ein direkter Vergleich von zwei oder mehr Intervallen andere statistische Eigenschaften aufweist als ein Test auf Gleichheit der Parameter. So führt beispielsweise der Vergleich zweier Parameter anhand der Überlappung von Konfidenzintervallen in der Regel zu konservativeren Entscheidungen als ein entsprechender Signifikanztest (Schenker & Gentleman, 2001). Der quasi visuelle Vergleich von Intervallen kann nicht ohne weiteres als Homogenitätstest interpretiert werden (vgl. Cumming & Finch, 2005). Der in Abbildung 1 skizzierte Ansatz wurde für Metaanalysen insbesondere im medizinischen Bereich zum sogenannten Forest Plot ausgearbeitet, bei welchem in horizontaler Ausrichtung die Punkt- und Intervallschätzungen der einzelnen Studien dem gemittelten Effekt graphisch gewichtet gegenübergestellt werden (Lewis & Clarke, 2001).

1.5 Untersuchungsgegenstand und Aufbau der Arbeit

Die bisherige Darstellung sollte deutlich machen, daß der Schätzung von Effektgrößen ein großer Stellenwert in der psychologischen Forschung zukommt und daß die statistischen Eigenschaften von Effektgrößen im Rahmen einer individuellen Studie, für den Vergleich mehrerer Untersuchungen sowie auch für die metaanalytische Zusammenfassung von Bedeutung sind. Die vorliegende Arbeit beschäftigt sich nun im wesentlichen mit den Eigenschaften zweier Effektgrößen, die für Unterschiedshypothesen bei zwei unverbundenen Stichproben entwickelt wurden, nämlich der standardisierten Lagedifferenz¹ sowie der stochastischen Überlegenheit $\Pr(X_1 > X_2)$.

Tabelle 1 gibt einen ersten Überblick über die in der Psychologie gebräuchlichen Datenmodelle, für welche die verschiedenen Lagetests entwickelt wurden und innerhalb derer die genannten Effektgrößen Verwendung finden. Bei dem parametrischen Modell (a) des Student-*t*-Tests wird von Normalverteilungen mit identischen Varianzen ausgegangen, ein zu testender Treatmenteffekt drückt sich demnach nur in einer Verschiebung der beiden Verteilungen aus. Im Normalverteilungsmodell (b) wird beschrieben, daß sich ein Treatmenteffekt sowohl in verschiedenen Mittelwerten als auch in verschiedenen Varianzen ausdrücken kann. Gefordert wird aber ebenfalls, daß ein Test nur auf die Differenz der Lokationen reagieren soll. Diese Testsituation wird als Behrens-Fisher-Problem bezeichnet, denn anders als für Modell (a) gibt es hierfür keinen gleichmäßig besten unverfälschten Test (Büning, 1991). Von den verschiedenen Lösungsvorschlägen für dieses Problem hat sich vor allem der Welch-Test als zufriedenstellend und praktikabel erwiesen.

Das Lokationsmodell (c) ist eine Verallgemeinerung des Modells (a), bei welchem davon ausgegangen wird, daß die Daten aus Verteilungen stammen, die aus der Verschiebung einer stetigen Verteilung F entstehen. Die Formen der beiden Verteilungen werden als identisch angenommen,

¹ Als standardisierte Lagedifferenz wird jede Statistik der Form $(\theta_1 - \theta_2)/\tau$ bezeichnet, wobei θ ein Lagemaß und τ ein Streuungsmaß sind. Die Begriffe Lage und Lokation bezeichnen allgemein die Position einer Verteilung auf einer Achse und werden hier synonym verwendet. Zur Anwendung kommen in der vorliegenden Arbeit das arithmetische Mittel, das 20%-getrimmte Mittel und das 20%-winsorisierte Mittel. Als Streuungs- bzw. Skalenmaße werden die herkömmliche Standardabweichung und die 20%-winsorisierte Streuung eingesetzt.

weshalb die hierfür entwickelten Tests auch als semi-parametrisch bezeichnet werden. Wie insbesondere Brunner und Munzel (2002) dargestellt haben, läßt sich das Funktional des U -Tests aber auch auf unstetige Verteilungen verallgemeinern. Die drei in Tabelle 1 genannten Tests gehören zusammen mit ihren Erweiterungen für den Mehrgruppenfall vermutlich zu den häufigst angewendeten inferenzstatistischen Verfahren in der Psychologie überhaupt. So wurden in 91% der 436 empirischen Arbeiten, die in den Jahren 2001 und 2002 in deutschsprachigen Zeitschriften erschienen, Mittelwerte mit t -Tests und Varianzanalysen ausgewertet und in immerhin noch 10% der Arbeiten Rangverfahren wie der U -Test eingesetzt (Hager, 2005).

Tabelle 1. Datenmodelle, Tests und Effektgrößen bei unverbundenen Stichproben

	Modell	Hypothese	Test	Effektgröße
(a)	Normalverteiltes Lokationsmodell $X_1 \sim N(\mu_1, \sigma^2)$ und $X_2 \sim N(\mu_2, \sigma^2)$	$H_0 : \mu_1 = \mu_2$	t -Test	Cohens d
(b)	Normalverteiltes Lokations-Skalenmodell $X_1 \sim N(\mu_1, \sigma_1^2)$ und $X_2 \sim N(\mu_2, \sigma_2^2)$	$H_0 : \mu_1 = \mu_2$	Welch-Test	Cohens d'
(c)	Lokationsmodell $X_1 \sim F(x - \mu_1)$ und $X_2 \sim F(x - \mu_2)$	$H_0 : F_1 = F_2$	U -Test	$\Pr(X_1 > X_2)$

Da die Kenntnis der den beobachteten Daten zugrundeliegenden Verteilungen in der Forschungspraxis eher die Ausnahme ist – eine solche kann in der Psychologie z.B. bei Testscores vorliegen –, wurde die Robustheit der parametrischen und nicht- bzw. semi-parametrischen Testverfahren in zahlreichen Simulationsstudien untersucht (Bünig, 1991; Wilcox, 2005). Aus diesen Robustheitsstudien zu Tests läßt sich aber in der Regel nur wenig zum Überdeckungsverhalten (Einhaltung des Konfidenzniveaus) der hierzu äquivalenten Intervallschätzer erfahren, denn nur bei Gültigkeit der Nullhypothese ist die empirische Überdeckungswahrscheinlichkeit gleich $1 - \hat{\alpha}^*$, wobei $\hat{\alpha}^*$ die simulierte, relative Häufigkeit des Fehlertyps I ist. Zum Überdeckungsverhalten bei Gültigkeit von Alternativhypothesen kann hingegen keine Aussage abgeleitet werden.

Die zentrale Fragestellung der vorliegenden Arbeit lautet nun: Wie gut gelingt die Schätzung von Effektgrößen bei Abweichungen von den Modellen, die den Schätzern zugrundeliegen? Analog zu den verschiedenen Strategien zur Entwicklung robuster Tests werden hier die aktuell diskutierten Verfahren untersucht, die eine bessere Einhaltung des nominalen Niveaus bei nichtnormalen oder varianzheterogenen Verteilungen versprechen, als dies von den klassischen Methoden zu erwarten ist. Zudem werden die für unverbundene Stichproben entwickelten Effektgrößen auch auf den Fall verbundener Stichproben übertragen und anhand korrelierter Verteilungen untersucht. Ziel ist es, zumindest auf statistischer Ebene die Vergleichbarkeit zwischen Designs mit unabhängigen und abhängigen² Gruppen zu ermöglichen (Becker, 1988), wobei hier das Problem besteht, daß z.B. die Verteilung von \hat{d} selbst im Normalverteilungsmodell bei abhängigen Stichproben nicht bekannt ist (Viechtbauer, 2007).

Die vorliegende Arbeit umfaßt insgesamt vier Simulationsstudien. In Kapitel 2.1 werden die verschiedenen Varianten der Effektgröße d im Falle unverbundener Stichproben untersucht, wobei eine Robustifizierung der Intervallschätzer über die Verwendung getrimmter Mittelwerte sowie die

² Die Begriffe *abhängige Stichprobe* und *verbundene Stichprobe* werden im folgenden synonym verwendet.

Schätzung anhand von nichtzentralen und Bootstrapverteilungen erfolgt. Für den abhängigen Fall werden in Kapitel 2.2 mit diesen Methoden neben der Effektgröße d_D aus dem t -Test für abhängige Stichproben auch die für den unabhängigen Fall vorgestellten Effektgrößen simuliert. Analog zu Kapitel 2 wird für die stochastische Überlegenheit zunächst der Fall unverbundener Stichproben (Kapitel 3.1) und schließlich der Fall verbundener Stichproben (Kapitel 3.2) untersucht, wobei sich bei Abhängigkeit ebenfalls zwei Varianten der stochastischen Überlegenheit betrachten lassen. Anders als für die standardisierten Lagedifferenzen entfällt in Kapitel 3 die Untersuchung der Punktschätzer, da diese generell erwartungstreu sind (Birnbaum, 1956; Cliff, 1996b). Als Beurteilungskriterien werden bei Punktschätzern der Bias und der mittlere quadratische Fehler, bei Intervallschätzern die simulierte Überdeckungswahrscheinlichkeit und die Breite (bzw. Länge) der Intervalle herangezogen. Bei der Darstellung und Interpretation der Ergebnisse wird jeweils die Ableitung von Empfehlungen für die Praxis angestrebt. Die Arbeit schließt nach einer Illustration ausgewählter Schätzer an zwei Datenbeispielen aus der Forschung (Kapitel 4) mit einer Zusammenfassung und Diskussion in Kapitel 5.

2 Standardisierte Lagedifferenzen

Maße der Effektgröße wurden vor allem auf zweierlei Art definiert, als standardisierte Lagedifferenz und als Anteil aufgeklärter Varianz. Nach Winer, Brown und Michels (1991) kommt der erste Ansatz vor allem in Untersuchungen mit festen Effekten, der zweite in Untersuchungen mit zufälligen Effekten zur Anwendung. Während Maße der Varianzaufklärung die Unterschiedlichkeit von zwei oder mehr Gruppen auszudrücken vermögen, sind standardisierte Lagedifferenzen im wesentlichen bei zwei Gruppen sinnvoll und können in Mehrgruppendesigns als multiple Paarvergleiche Verwendung finden. Da die Faktoren in Zweigruppendesigns in der Regel fest sind, spielt die Varianzaufklärung hier eine untergeordnete Rolle.

2.1 Unverbundene Stichproben

Im folgenden werden die gängigen Definitionen der Parameter sowie verschiedene verzerrte und korrigierte Punktschätzer vorgestellt. Im Anschluß werden Möglichkeiten der Intervallschätzung aufgezeigt. Da manche dieser Verfahren wie z.B. die Verbindung von Bootstrap mit robustifizierten Schätzern in der Psychologie weniger üblich sind, werden diese zuerst vorgestellt, bevor eine Begründung für die in der Simulationsstudie berücksichtigten Verfahren gegeben wird.

2.1.1 Parameter und Punktschätzer

2.1.1.1 Cohens d und d'

Das bekannteste und meist eingesetzte Maß der Effektgröße ist die erwähnte standardisierte Mittelwertsdifferenz nach Cohen (1988, S. 20):

$$d = \frac{\mu_1 - \mu_2}{\sigma},$$

wobei σ die den beiden Populationen gemeinsame Streuung bezeichnet. Die Effektgröße d gibt also an, um wieviele Standardabweichungen die Erwartungswerte von zwei varianzhomogenen Verteilungen voneinander abweichen. Die lateinische Notation stammt von Cohen und wird hier beibehalten, obwohl es sich bei d um einen (Populations-)Parameter handelt. Zur Berechnung der Effektgröße auf Ebene der Stichprobe schlägt Cohen die gleiche Statistik vor, bei der die Verteilungsparameter jeweils durch erwartungstreue Schätzer ersetzt werden:

$$d_s = \frac{\bar{X}_1 - \bar{X}_2}{S_p}, \text{ wobei}$$

$$S_p = \sqrt{\frac{\sum (X_{1i} - \bar{X}_1)^2 + \sum (X_{2j} - \bar{X}_2)^2}{n_1 + n_2 - 2}}.$$

Zum besseren Verständnis der Effektgröße d bietet Cohen (1988, S. 21 f.) die folgende Beziehung zur punkt-biseriellen Korrelation an: $\rho_{pb} = d / \sqrt{d^2 + 4}$. Die Effektgröße d kann also zum einen in ein normiertes Korrelationsmaß und nach Quadrieren auch in ein Maß der aufgeklärten Varianz überführt werden. Zum anderen bietet er mit den U -Maßen einen Ansatz an, d in Flächenanteilen von Normalverteilungen auszudrücken. Sofern die Populationen gemäß dem Modell des t -

Tests mit gleichen Varianzen normalverteilt sind, kann d in folgende Anteile der (Nicht-)Überlappung überführt werden:

$$U_3 = \Phi(|d|),$$

$$U_2 = \Phi\left(\frac{|d|}{2}\right),$$

$$U_1 = \Phi\left(\frac{2\Phi\left(\frac{|d|}{2}\right) - 1}{\Phi\left(\frac{|d|}{2}\right)}\right) = \frac{2U_2 - 1}{U_2},$$

wobei $\Phi(\cdot)$ die Standardnormalverteilungsfunktion bezeichnet. Da Cohen die U -Maße nur für $d \geq 0$ diskutiert und insbesondere U_1 auch nur für diesen Wertebereich sinnvoll ist, wurden die Formeln hier um Betragstriche ergänzt. Abbildung 2 stellt die von den U -Maßen erfaßten Flächen dar.

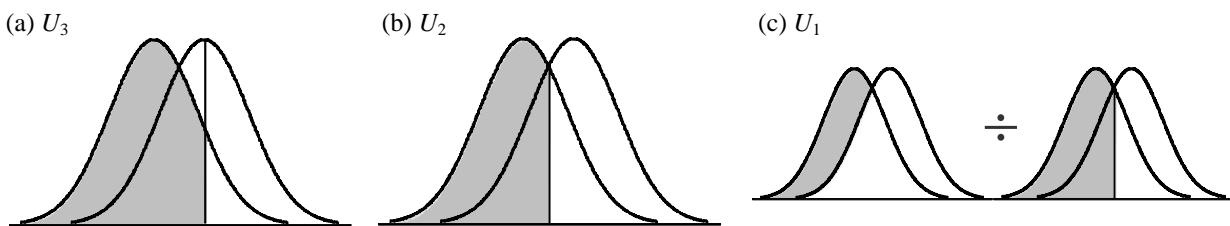


Abbildung 2. U -Maße zur Interpretation der Effektgröße d

U_3 gibt den Anteil einer der beiden Populationen wieder, der oberhalb bzw. unterhalb des Mittelwertes der jeweils anderen Population liegt. U_2 entspricht der Hälfte der Gesamtfläche beider Populationen – mit Gesamtfläche ist $\int \max[f_1(x), f_2(x)]dx$ gemeint –, somit gibt U_2 den Anteil einer Population an, der dem gleichgroßen Anteil der jeweils anderen Population entgegengesetzt ist. U_1 gibt schließlich den Anteil der nicht überlappenden Fläche an der Gesamtfläche wieder, wobei dies gleichermaßen für beide Populationen wie auch für das gesamte Modell gilt. Im Falle einer Therapiestudie beispielsweise werde angenommen, Therapie B zeitige im Vergleich zu Therapie A ein höheres Wohlbefinden. Die Hypothese $H_1: d = 0.5$ übersetzt in Einheiten der U -Maße bedeutet dann, (U_3) daß sich 69% der Patienten aus Therapie B besser fühlen als die Hälfte der Patienten aus Therapie A, daß (U_2) 60% der Patienten aus Therapie B ein höheres Wohlbefinden haben als 60% der Patienten aus Therapie A, und schließlich (U_1) daß sich das Wohlbefinden von 33% aller Patienten nicht mit der jeweils anderen Therapie erreichen oder erklären läßt.

Die Empfehlung, Mittelwertsvergleiche möglichst auch um ein Maß der Verteilungsüberlappung zu ergänzen, geht auf Kelley (1920) und Tilton (1937) zurück. Anders als Cohen betrachten diese Autoren aber nicht die Nichtüberlappung, sondern die gemeinsame Fläche (d.i. $\int \min[f_1(x), f_2(x)]dx$) zweier varianzhomogener Normalverteilungen und definieren hierfür folgende Statistik:

$$O_v = 2 \cdot \Phi\left(-\frac{|\mu_1 - \mu_2|}{2\sigma}\right) = 2 \cdot \Phi\left(-\frac{|d|}{2}\right).$$

In Tabelle 2 sind nun die Beziehungen zwischen den dargestellten Maßen anhand einiger Werte aufgelistet.

Tabelle 2. Wertevergleich zwischen d , ρ_{pb} und den U -Statistiken sowie Tiltons Überlappungsmaß

$ d $	U_3	U_2	U_1	O_v	ρ_{pb}	ρ_{pb}^2
0	0.500	0.500	0	1	0	0
0.2	0.579	0.540	0.148	0.920	0.100	0.010
0.5	0.691	0.599	0.330	0.803	0.243	0.059
0.8	0.788	0.655	0.474	0.689	0.371	0.138
1.0	0.841	0.691	0.553	0.617	0.447	0.200

Anders als die U -Maße oder andere Flächeninterpretationen wie Tiltons Überlappungsmaß³, die in der Praxis kaum eingesetzt werden, hat sich vor allem Cohens sprachliche Kategorisierung der Werte von Effektgrößen in der psychologischen Forschung durchgesetzt. Demnach werden Effekte um $d = 0.2$ als klein, um $d = 0.5$ als mittel und Effekte um $d = 0.8$ als groß bezeichnet. Zielsetzung dieser Klassifizierung der Effektgrößen ist es, die Durchführung von Teststärkeanalysen auch dann zu ermöglichen, wenn nur mangelnde Vorinformationen über die Verteilung der zu erhebenden abhängigen Variablen vorliegen und das statistische Modell zu unspezifisch ist, um eine Effektgröße ableiten zu können. In diesem Fall gelingt eine Optimierung des Signifikanztests nur, wenn eine Effektgröße unmittelbar festgelegt wird. Diesen kritischen Richtwerten, die für die Effektgrößen aller parametrischen Tests festgelegt wurden, liegt sowohl eine eher theoretische als auch eine empirische Argumentation zugrunde (Cohen, 1962; 1988). So wird beispielsweise das Konzept eines mittleren Effekts dadurch verdeutlicht, daß ein solcher Effekt gerade so groß ist, um mit bloßem Auge entdeckt werden zu können (Cohen, 1988, S. 26). Als empirische Befunde mit mittlerem Effekt werden der Größenunterschied von 14 zu 18 jährigen Mädchen oder auch die Differenz der mittleren Testintelligenz von Büroangestellten und angelernten Arbeitern angeführt. Cohen selbst hat wiederholt auf die unvermeidliche Willkür bei der Festlegung sowie den nur relativen Wert solcher Konventionen hingewiesen. Relativ sind hierbei nicht nur die Begriffe zueinander, sondern auch ihre Bedeutung im Hinblick auf verschiedene Forschungsfelder und Methoden.

Für den Fall ungleicher Varianzen definiert Cohen die Effektgröße analog zu d , jedoch an Stelle von σ unter Verwendung der gemittelten Varianzen:

$$d' = \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}}.$$

Als Schätzer dient $d'_s = (\bar{X}_1 - \bar{X}_2) / \sqrt{\frac{1}{2}(S_1^2 + S_2^2)}$. Eine flächenorientierte Interpretation von d' ist nicht ohne weiteres möglich, denn die Überlappung hängt sowohl von den Mittelwerts- als auch

³ Neuere Schätzer der Überlappung zweier beliebiger Verteilungen wurden von Wilcox und Muska (1999) sowie von Huberty und Lowman (2000) auf Basis der Güte (Hitrate) von Klassifikationsregeln entwickelt.

von den Varianzunterschieden ab. In Abbildung 3 wird die Überlappung zweier Normalverteilungen in Abhängigkeit von der Effektgröße d' sowie dem Verhältnis der beiden Varianzen dargestellt.

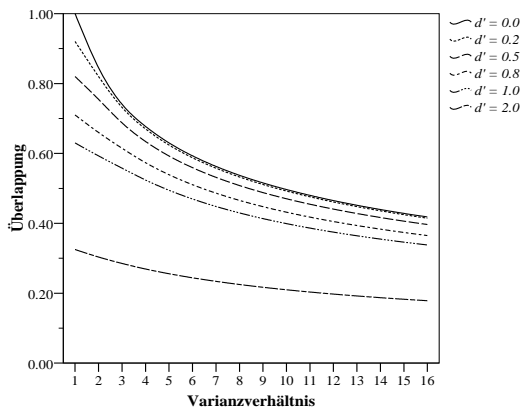


Abbildung 3. Überlappung zweier Normalverteilungen in Abhängigkeit von Effektgröße d' und Varianzverhältnis $\sigma_1^2 : \sigma_2^2$

Bei konstant gehaltener Effektgröße d' nimmt die Überlappung der Normalverteilungen also mit zunehmender Varianzheterogenität ab. d' hat somit keinen konkreten Bezug zu den beiden Populationsverteilungen, da die Standardisierung an einer anderen, aggregierten Einheit erfolgt. Neben der mangelnden Robustheit (vgl. Abbildung 4 in Kapitel 2.1.1.6) ist diese fragliche Interpretierbarkeit von d' bei Varianzheterogenität für Wilcox (1996) ein wesentliches Argument gegen die Verwendung von standardisierten Lagedifferenzen zur Beschreibung eines Effekts. Denn in der Praxis ist nach Wilcox in aller Regel davon auszugehen, daß sich Treatments sowohl auf die Lage als auch auf die Variabilität der Daten auswirken. Für eine Teststärkeanalyse stellt die Varianzheterogenität hingegen kein Problem dar, d.h. die Fehlerwahrscheinlichkeiten lassen sich insbesondere bei gleich großen und nicht zu kleinen Gruppen korrekt kalkulieren. Auch die genannte Beziehung zur punktbiseriellen Korrelation gilt weiterhin.

2.1.1.2 Glass' Δ

Glass (1976) führte unabhängig von Cohen die Effektgröße Δ ein, die mit d jedoch identisch ist, weshalb die Bezeichnung nach Glass hier nicht benutzt wird. Der wesentliche Beitrag von Glass liegt in der Behandlung der Effektgröße als Zufallsvariable, die über eine Vielzahl von Studien betrachtet zur Einschätzung der Konsistenz und Größe eines Phänomens herangezogen werden kann (Glass, McGaw & Smith, 1981). Während Cohen die Zusammenhänge zwischen Effektgröße und *Testen* in der primären Forschung behandelt, führt Glass in die Problematik des *Schätzens* von Effektgrößen im Rahmen der Metaanalyse ein.

Glass geht ebenfalls von einem Modell mit gleichen Populationsvarianzen aus und nimmt die Möglichkeit ernst, daß in der Effektgröße allein der Lageunterschied zum Ausdruck kommen soll und kann. Da sich beim Vergleich *mehrerer* Gruppen aber das Problem ergeben kann, daß der deskriptiv kleinere Mittelwertsunterschied zu einer größeren Effektgrößenschätzung führt, weil die Varianzschätzung zufallsbedingt klein ausfällt, schlägt Glass vor, die Mittelwertsdifferenzen mehrerer Gruppen immer an derselben Streuung zu relativieren, wobei er die Streuung der Kontrollgruppe wählt: $d_s^{S_1} = (\bar{X}_1 - \bar{X}_2) / S_1$, Gruppe 1 bezeichne hier also die Kontrollgruppe. Für den von Glass

betrachteten homoskedastischen Fall ist $d_s^{S_1}$ auch ein Schätzer für d . Dieses Vorgehen wurde hingegen von Hedges und Olkin (1985, S. 78) abgelehnt, da die Bündelung aller Gruppenvarianzen zu einer präziseren Varianz- und also auch Effektgrößenschätzung führt. Da in der vorliegenden Arbeit die Eigenschaften der Schätzer auch für den heteroskedastischen Fall untersucht werden, wird $d_s^{S_k}$ (mit k als Index für die Gruppe) hier als Schätzer des entsprechend an nur einer Streuung standardisierten Parameters betrachtet:

$$d^{S_k} = \frac{\mu_1 - \mu_2}{\sigma_k}.$$

2.1.1.3 Hedges' erwartungstreuer Schätzer für d

Hedges (1981) konnte anhand der Funktion der nichtzentralen t -Verteilung zeigen, daß d_s ein positiv verzerrter Schätzer für d ist und der Erwartungswert wie folgt lautet:

$$E(d_s) = \frac{d}{c(m)},$$

$$\text{wobei } c(m) = \frac{\Gamma\left(\frac{m}{2}\right)}{\sqrt{m/2} \Gamma\left(\frac{(m-1)}{2}\right)}$$

und $m = n_1 + n_2 - 2$ oder, falls die Streuung auf Basis nur einer Gruppe geschätzt wird, $m = n_k - 1$. Γ bezeichnet die Gammafunktion, die eine stetige Verallgemeinerung der Fakultät ist (Krengel, 2003); für natürliche Zahlen gilt $\Gamma(x) = (x - 1)!$. Dieser Korrekturfaktor ist bei $c(2) = .564$ und geht mit zunehmendem Stichprobenumfang gegen Eins: $c(3) = .724$, $c(5) = .841$, $c(10) = .923$, $c(20) = .962$, $c(30) = .975$, $c(50) = .985$. Als erwartungstreuer Schätzer für d ergibt sich:

$$\hat{d} = c(m)d_s.$$

Da der Korrekturfaktor stets kleiner Eins ist, weist \hat{d} auch eine kleinere Varianz auf als d_s .

2.1.1.4 Huynhs Biaskorrektur bei Varianzheterogenität

Für den Fall heterogener Varianzen hat Huynh (1989) folgende Modifikation der Korrektur nach Hedges entwickelt, die im Vergleich zu d'_s zu einer weniger verzerrten und präziseren Schätzung von d' führt:

$$\hat{d}' = c(\hat{m}) \cdot d'_s, \text{ wobei}$$

$$\hat{m} = \frac{(n_1 - 1) \cdot (n_2 - 1) \cdot (S_1^2 + S_2^2)^2}{(n_1 - 1) \cdot S_1^4 + (n_2 - 1) \cdot S_2^4}.$$

Zu beachten ist, daß analog zu den Freiheitsgraden des Welch-Tests auch die Bias-Korrektur nach Huynh nur eine Schätzung der exakten, von den unbekanntem Populationsvarianzen abhängigen Korrektur ist. Die Eigenschaften dieser Korrektur wurden m.W. seit der originären Arbeit von Huynh nicht mehr untersucht.

2.1.1.5 Bootstrap-Biaskorrektur

Als genereller Ansatz zur Korrektur verzerrter Schätzer bietet sich das Bootstrap-Verfahren an, welches von Efron (1979) eingeführt wurde und zusammen mit Jackknife- und Permutationsverfahren zu den Stichprobenwiederholungstechniken gehört, mit denen durch wiederholte Ziehung von Stichproben (Resamples) aus einer vorliegenden Stichprobe die Verteilungsfunktion eines Schätzers auf empirischem Weg gewonnen wird. Die Technik des Jackknife-Verfahrens wird bei der Darstellung des BCA-Bootstraps kurz erläutert, auf Permutations- bzw. Randomisationsverfahren wird hier nicht weiter eingegangen.

Das wesentliche Kennzeichen des Bootstrap-Verfahrens ist die Ziehung von Stichproben mit Zurücklegen. Gegeben sei eine konkrete Zufallsstichprobe $X = (x_1, x_2, \dots, x_n)$, deren Komponenten $X_i \stackrel{iid}{\sim} F(x)$ sind; ferner sei $\hat{\theta}$ ein von X und $F(x)$ abhängender Schätzer für einen Parameter von $F(x)$, dessen Verteilungsfunktion nicht bekannt ist. Eine Zufallsstichprobe $X^* = (x_1^*, x_2^*, \dots, x_n^*)$, die mit Zurücklegen aus X gezogen wird, wird Bootstrap-Stichprobe und der Schätzer für θ auf Basis von X^* wird Bootstrap-Schätzer $\hat{\theta}^*$ genannt. Werden nun sämtliche möglichen Bootstrap-Stichproben aus einer konkreten Stichprobe gezogen und jeweils der Bootstrap-Schätzer eines Parameters bestimmt, so läßt sich durch Ordnung dieser Schätzwerte und Kumulation der jeweiligen Ziehungswahrscheinlichkeiten die exakte Bootstrap-Verteilung bestimmen. Diese empirisch gewonnene Verteilungsfunktion ist die Bootstrap-Schätzung der unbekanntem Verteilungsfunktion von $\hat{\theta}$. Da es aber bereits bei kleinen Stichproben zu aufwendig ist, alle möglichen unterschiedlichen Bootstrap-Stichproben zu berechnen – im Ein-Gruppen-Fall sind dies bereits $(2n-1)!/[(n-1)!n!]$ –, begnügt man sich in der Regel mit einer Approximation der exakten Bootstrap-Verteilung auf Basis von B -vielen Stichproben, wobei B je nach Anwendung und Bootstrap-Variante zwischen wenigen Hundert und mehreren Tausend liegt.

Aus einer gegebenen Stichprobe der Größe n werden B Zufallsstichproben vom Umfang n mit Zurücklegen gezogen und in jeder dieser Bootstrap-Stichproben wird der interessierende Schätzer $\hat{\theta}^*$ berechnet. Die Bootstrap-Biasschätzung ist nun die Differenz zwischen dem Mittelwert der Bootstrap-Verteilung und der Schätzung aus der konkreten Stichprobe. Der korrigierte Schätzer lautet dann:

$$\hat{\theta}^{Boot} = \hat{\theta} - Bias^{Boot} = \hat{\theta} - (\hat{\theta}_{(\cdot)}^* - \hat{\theta}), \text{ wobei } \hat{\theta}_{(\cdot)}^* = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b^*.$$

Diese Bias-Korrektur läßt sich insofern als nichtparametrisch bezeichnen, als sie allein aus der empirischen Verteilungsschätzung (konkrete Stichprobe) abgeleitet wird, ohne daß theoretische Annahmen über die Populations- oder die Stichprobenverteilung gemacht werden.

2.1.1.6 Effektgrößen mit getrimmten Mittelwerten

Während in der klassischen parametrischen Statistik von a priori bekannten Dichtefunktionen ausgegangen wird und die Optimalität eines Schätzers oder Tests sich auf genau ein Verteilungs-

modell beschränkt, wird in der robusten Statistik nach Verfahren gesucht, die ihre Effizienz möglichst beibehalten, wenn von dem Modell abgewichen wird, für welches sie optimale Eigenschaften besitzen (Büning, 1991). Obwohl die Normalverteilung vielfach eine gute Approximation anderer Verteilungen darstellt, verlieren die unter ihr entwickelten kleinsten Quadrat-(KQ-)Schätzer auch bei kleinen Abweichungen von diesem Modell schnell ihre Güte. Aus theoretischer und auch praktischer Perspektive werden vor allem Asymmetrie und starke Tails⁴ von Verteilungen sowie das Auftreten von Ausreißern als problematisch für die Anwendung der klassischen parametrischen Statistik angesehen. Von der Vielzahl robustifizierter Lagemaße (Andrews, Bickel, Hampel, Huber, Rogers & Tukey, 1972) hat sich in der psychologischen Literatur am ehesten das getrimmte Mittel durchgesetzt (Wilcox, 2005). Dies ist zum einen mit relativ guten Schätzeigenschaften über einen weiten Bereich von Verteilungen hinweg zu begründen, zum anderen liegt mit der winsorisierten Varianz ein einfacher Schätzer des Standardfehlers des getrimmten Mittels vor, mit dem sich robustifizierte t -Tests entwickeln lassen (Tukey & McLaughlin, 1963; Yuen, 1974; Yuen & Dixon, 1973). Analog zur Definition der Effektgrößen d und d' in den Modellen des klassischen und des Welch- t -Tests haben Hogarty und Kromrey (2001) sowie Algina, Keselman und Penfield (2005a) nun Alternativen mit getrimmten Mittelwerten vorgeschlagen, die den Modellen der robustifizierten t -Tests folgen. Formal lautet die von Algina et al. vorgeschlagene Alternative für d :

$$d_R^{\text{Algina}} = 0.642 \frac{\mu_{t1} - \mu_{t2}}{\sigma_w},$$

wobei die Parameter mit jeweils 20% getrimmt bzw. winsorisiert sind. Die Wahl dieses Trimmanteils geht wesentlich auf die Arbeiten von Wilcox (1996, 2005) zurück. Aufgrund der Multiplikation mit 0.642 – das ist der Parameterwert der 20%-winsorisierten Streuung der Standardnormalverteilung – gilt zur besseren Vergleichbarkeit mit der herkömmlichen Effektgröße $d = d_R^{\text{Algina}}$ für Normalverteilungen mit gleichen Varianzen. Als Schätzer für diese robustifizierte Effektgröße dient

$$d_{s,R}^{\text{Algina}} = 0.642 \frac{\bar{X}_{t1} - \bar{X}_{t2}}{S_{(w)p}}, \text{ wobei}$$

$$S_{(w)p} = \sqrt{\frac{(n_1 - 1)S_{w1}^2 + (n_2 - 1)S_{w2}^2}{n_1 + n_2 - 2}}$$

die gebündelte 20%-winsorisierte Streuung ist. Um Sprünge im tatsächlichen Trimmanteil zu vermeiden, die bei ganzzahligem Trimmen entstehen, wird hier das geglättete, symmetrisch getrimmte Mittel (Rosenberger & Gasko, 1983, S. 311) betrachtet:

⁴ Während viele Wahrscheinlichkeitsdichten die Eigenschaft haben, daß $f(x)$ mit $x \rightarrow \pm \infty$ schnell gegen Null konvergiert, besitzen Verteilungen mit starken bzw. langen Tails vergleichsweise große Wahrscheinlichkeitsmasse an den Rändern. Da weder die Streuung noch die Wölbung zur Beschreibung dieser Eigenschaft geeignet sind, wurden verschiedene Tailmaße entwickelt, die vor allem im Rahmen des adaptiven Testens von Bedeutung sind (Büning, 1991).

$$\bar{X}_{\gamma} = \frac{1}{n(1-2\gamma)} \left\{ (1-q)[X_{(g+1)} + X_{(n-g)}] + \sum_{i=g+2}^{n-g-1} X_{(i)} \right\},$$

wobei der Trimmanteil allgemein $0 \leq \gamma < .5$, hier jedoch stets $.2$, für jede Seite der geordneten Statistik $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$, $q = \gamma n - g$ und $g = [\gamma n]$ sind; $[\gamma n]$ ist die auf die nächste natürliche Zahl abgerundete Anzahl γn . Auf eine Indizierung der Gruppe wurde hier verzichtet. Beim symmetrischen Trimmen wird also auf beiden Seiten der gleiche Prozentsatz der extremsten Werte von der weiteren Berechnung ausgeschlossen. Beim Winsorisieren hingegen wird die Anzahl der Werte nicht reduziert, sondern die durch den gewählten Prozentsatz erfaßten Randwerte werden durch den jeweils äußersten nicht erfaßten Wert ersetzt. Analog wird auch bei den winsorisierten Statistiken zwischen auszuschließenden und verbleibenden Werten gemittelt, sofern γn keine ganze Zahl ist. Das winsorisierte Mittel lautet dann:

$$\bar{X}_{\gamma w} = \frac{1}{n} \left\{ (g+1)[(1-q)(X_{(g+1)} + X_{(n-g)}) + q(X_{(g+2)} + X_{(n-g-1)})] + \sum_{i=g+2}^{n-g-1} X_{(i)} \right\}.$$

Die winsorisierte Varianz schließlich kann wie folgt dargestellt werden:

$$S_{\gamma w}^2 = \frac{1}{n-1} \left\{ (g+1) \left\{ \begin{aligned} & \left[(1-q)X_{(g+1)} + qX_{(g+2)} - \bar{X}_{\gamma w} \right]^2 \\ & + \left[(1-q)X_{(n-g)} + qX_{(n-g-1)} - \bar{X}_{\gamma w} \right]^2 \end{aligned} \right\} + \sum_{i=g+2}^{n-g-1} (X_{(i)} - \bar{X}_{\gamma w})^2 \right\}.$$

Die robustifizierten Varianten der anderen Effektgrößen lauten entsprechend:

$$d_R^{\text{Algina}} = 0.642 \frac{\mu_{t1} - \mu_{t2}}{\sqrt{\frac{\sigma_{w1}^2 + \sigma_{w2}^2}{2}}} \text{ und}$$

$$d_R^{S_i \text{ Algina}} = 0.642 \frac{\mu_{t1} - \mu_{t2}}{\sigma_{wk}}.$$

Die Berechnung der robustifizierten Schätzer sei in Tabelle 3 an einem Beispiel mit $n_1 = 13$ und $n_2 = 9$ illustriert. Für das 20%-getrimmte Mittel von X_1 müßten an jeder Seite $.2 \cdot 13 = 2.6$ Werte abgeschnitten werden; daher gehen die Werte $X_{1(4)}$ bis $X_{1(10)}$ mit vollem Gewicht von Eins und die Werte $X_{1(3)}$ und $X_{1(11)}$ jeweils mit einem Gewicht von $.4$ in die Summenbildung ein. Die Summe der Gewichte ist damit gleich der Anzahl zu mittelnder Werte: $n_1(1 - 2 \cdot .2) = 7.8$. Bei der Winsorisierung werden in dieser Gruppe jeweils die drei äußeren Werte auf beiden Seiten auf einen analog gewichteten Wert gesetzt. Die unteren winsorisierten Werte ergeben sich aus $.4 \cdot X_{1(3)} + .6 \cdot X_{1(4)} = 1434.8$, die oberen Werte lauten $.6 \cdot X_{1(10)} + .4 \cdot X_{1(11)} = 2049.4$. Die winsorisierte Varianz ergibt sich nun anhand der üblichen, mit diesen winsorisierten Werten versehenen Formel. Wie die Werte in den Formeln der Effektgrößen in Tabelle 3 verdeutlichen, fällt insbesondere die winsorisierte Varianz in Gruppe 1 ($S_{w1}^2 = 71992.930$) deutlich kleiner aus als die herkömmliche Varianz ($S_1^2 = 1992338.231$). Zudem liegt das getrimmte Mittel in dieser Gruppe ($\bar{X}_{t1} = 1684.128$) unterhalb des üblichen Mittelwertes ($\bar{X}_1 = 2131.308$). Ursache hierfür sind die beiden

Ausreißerwerte auf der rechten Seite der Gruppe 1. Diese führen im Zusammenhang mit den anderen Werten zu einem erheblichen Anstieg der Varianz, weshalb d'_s deutlich kleiner ausfällt als die robustifizierte Variante. Zu beachten ist, daß die in der Praxis anzutreffende (automatische) Bereinigung von Daten um solche Extrem- oder Ausreißerwerte vor der Verwendung von KQ-Statistiken problematisch ist, sofern diese Werte nicht eindeutig als falsche oder unmögliche Werte identifiziert werden können. Wenn es sich nämlich um echte Beobachtungen handelt, wird der Standardfehler des bereinigten Mittelwertes anhand der herkömmlichen Varianz unterschätzt (Wilcox, 1998a).

Tabelle 3. Beispiel zur Berechnung der robustifizierten Effektgröße

(a) Beobachtete Werte													
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
X_1	1250	1353	1382	1470	1480	1567	1632	1641	1934	2025	2086	3456	6431
X_2	1109	1110	1140	1249	1314	1358	1388	1424	1424				
(b) 20%-getrimmte Werte													
X_{t1}			(552.8)	1470	1480	1567	1632	1641	1934	2025	(834.4)		
X_{t2}		(222)	1140	1249	1314	1358	1388	(284.8)					
(c) 20%-winsorisierte Werte													
X_{w1}	1434.8	1434.8	1434.8	1470	1480	1567	1632	1641	1934	2025	2049.4	2049.4	2049.4
X_{w2}	1134	1134	1140	1249	1314	1358	1388	1395	1395				

$$d'_{s,R}{}^{\text{Algina}} = 0.642 \frac{\bar{X}_{t1} - \bar{X}_{t2}}{\sqrt{\frac{S_{w1}^2 + S_{w2}^2}{2}}} = 0.642 \frac{1684.128 - 1288.111}{\sqrt{\frac{71992.930 + 13577.680}{2}}} = 1.229$$

$$d'_s = \frac{2131.308 - 1279.555}{\sqrt{\frac{1992338.231 + 17384.528}{2}}} = 0.850$$

Vorteile (hinsichtlich der Resistenz) robustifizierter Effektgrößen lassen sich mit ihrem Verhalten unter Verteilungen in der Nähe von der Normalverteilung begründen, bei denen die klassischen Effektgrößen aufgrund ihrer Anfälligkeit für Extremwerte zusammenbrechen. Abbildung 4a zeigt zwei Normalverteilungen mit Varianz 1, deren Dichten um 0.8 Skaleneinheiten verschoben sind, so daß sich als Effektgröße $d = 0.8$ ergibt. In Abbildung 4b sind zwei kontaminierte Normalverteilungen dargestellt, die ebenfalls um 0.8 Skaleneinheiten verschoben sind. Eine kontaminierte Normalverteilung stellt eine Mischung von Normalverteilungen dar, wobei hier die Werte mit Wahrscheinlichkeit .9 aus einer Normalverteilung mit einer Streuung von 1 und mit Wahrscheinlichkeit .1 aus einer Normalverteilung mit einer Streuung von 10 gezogen wurden. Die so kontaminierte Verteilung hat eine Streuung von 3.3. Somit ergibt sich für dieses Modell eine Effektgröße von nur noch $d = 0.24$. Nicht nur visuell sind die Abbildungen 4a und 4b recht ähnlich, sondern auch im Hinblick auf ihre Verteilungsdistanzen lassen sie sich als ähnlich bezeichnen (Gleason, 1993; Wilcox, 1998a). d_R^{Algina} beschreibt die Unterschiedlichkeit der Verteilungen in Abbildung 4a ebenfalls mit

0.8 und die Situation in Abbildung 4b mit 0.78, was der offensichtlichen Ähnlichkeit der Datensituationen eher gerecht wird. Analog hierzu würde die Betrachtung der korrespondierenden Testverfahren ergeben, daß der klassische t -Test aufgrund der hohen Varianz unter den kontaminierten Verteilungen eine deutlich geringere Stärke besitzt als der getrimmte Test. Zum Vergleich zeigt Abbildung 4c zwei Normalverteilungen mit $d = 0.24$.

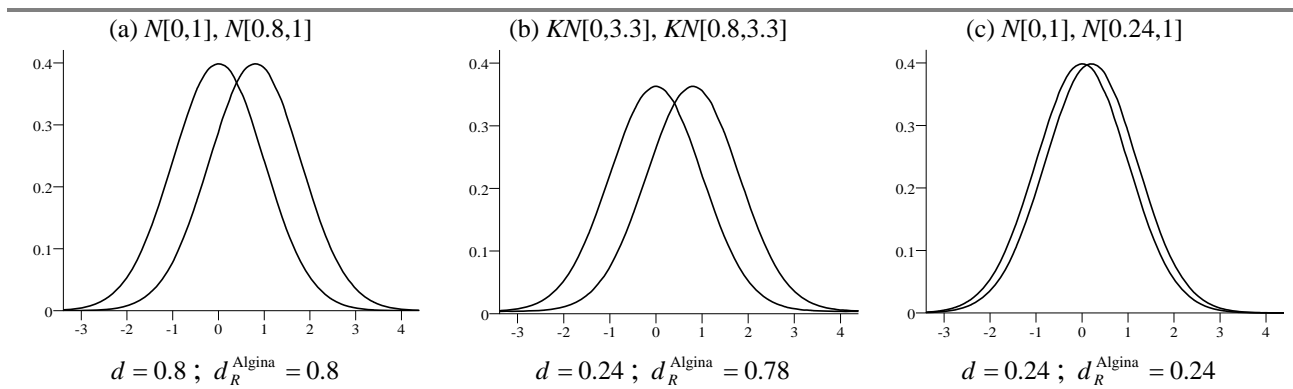


Abbildung 4. Effektgrößen bei Normalverteilungen und kontaminierten Normalverteilungen

2.1.2 Intervallschätzer

In der psychologischen Methodenlehre werden typischerweise Konfidenzintervalle unterrichtet und in der Praxis eingesetzt, die auf studentisierten Statistiken der Form $(\hat{\theta} - \theta)/\hat{\sigma}_{\hat{\theta}}$ beruhen. Wenn die Verteilung einer studentisierten Statistik unabhängig von dem zu schätzenden Parameter θ ist, welche dann auch als Pivot bezeichnet wird (Rohatgi, 1984, S. 694), kann ein (zentrales) Konfidenzintervall wie folgt konstruiert werden: $P(\hat{\theta} - c_o \cdot \hat{\sigma}_{\hat{\theta}} \leq \theta \leq \hat{\theta} - c_u \cdot \hat{\sigma}_{\hat{\theta}}) = 1 - \alpha$, wobei $\hat{\sigma}_{\hat{\theta}}$ der geschätzte Standardfehler von $\hat{\theta}$, c_u und c_o der untere und obere kritische Wert der studentisierten Verteilung und $1 - \alpha$ das Konfidenzniveau sind. Vielfach aber ist die Verteilung der studentisierten Statistik nur für $n \rightarrow \infty$ bekannt, so daß viele der Standardverfahren nur asymptotische Eigenschaften haben und erst für größere Stichproben das Konfidenzniveau hinreichend genau einhalten (Haagen & Seifert, 1979; Sievers, 1998).

2.1.2.1 Zentrale Schätzer

Auch für die Effektgröße d wurden asymptotische Verfahren vorgeschlagen, weil der Rechenaufwand für das nichtzentrale Vorgehen unverhältnismäßig hoch erschien. Da die Verteilung von \hat{d} mit zunehmender Stichprobengröße in eine Normalverteilung übergeht (Johnson & Welch, 1939), können die Grenzen eines Konfidenzintervalls über

$$d_u^\infty = \hat{d} - z_{1-\alpha/2} \hat{\sigma}_{\hat{d}}, \quad d_o^\infty = \hat{d} - z_{\alpha/2} \hat{\sigma}_{\hat{d}}$$

bestimmt werden, wobei $\hat{\sigma}_{\hat{d}} = \{(n_1 + n_2)/(n_1 n_2) + \hat{d}^2/[2(n_1 + n_2)]\}^{0.5}$ die asymptotische Streuung von \hat{d} und $z_{(\cdot)}$ kritische Werte der Standardnormalverteilung sind (Hedges & Olkin, 1985, S. 86). Auf die weiteren, ebenfalls bei Hedges und Olkin (S. 85 ff.) dargestellten asymptotischen Intervalle sei hier nur hingewiesen, da diese keine günstigeren Eigenschaften aufweisen als das nun folgende Intervall auf Basis der nichtzentralen t -Verteilung.

2.1.2.2 Nichtzentrale Schätzer

Bei der Berechnung des nichtzentralen Intervalls wird zunächst ein Konfidenzintervall für den Nichtzentralitätsparameter der t -Verteilung bestimmt, dessen Grenzen dann in die Einheit der Effektgröße transformiert werden (Steiger & Fouladi, 1997; Venables, 1975). Im Modell des t -Tests folgt die Teststatistik $t_{emp} = (\bar{X}_1 - \bar{X}_2) / (S_p \sqrt{1/n_1 + 1/n_2})$ einer nichtzentralen t -Verteilung mit $v = n_1 + n_2 - 2$ Freiheitsgraden und Nichtzentralitätsparameter $\delta = d \cdot \sqrt{(n_1 \cdot n_2) / (n_1 + n_2)}$. Nur bei Gültigkeit der Nullhypothese ist $\delta = 0$ und t_{emp} hat die zentrale, um Null symmetrische Verteilung, wie sie in Auszügen in Lehrbüchern tabelliert ist. Um nun ein Konfidenzintervall für δ angeben zu können, müssen für einen beobachteten Wert t_{emp} die Nichtzentralitätsparameter δ_u und δ_o gefunden werden, für die gilt:

$$P(t_{v, \delta_u} \geq t_{emp}) = \alpha/2, \quad P(t_{v, \delta_o} \leq t_{emp}) = \alpha/2.$$

Die Konfidenzgrenzen δ_u und δ_o können jedoch nicht über eine geschlossene Formel berechnet werden, sondern müssen in der Regel über iterative Algorithmen computergestützt bestimmt werden. Schließlich werden δ_u und δ_o über die obige Beziehung zur Effektgröße in die interessierenden Konfidenzgrenzen umgerechnet:

$$d_u = \delta_u \cdot \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}}, \quad d_o = \delta_o \cdot \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}}.$$

Sofern das Modell des t -Tests gegeben ist, ist das so konstruierte Intervall ein exaktes Konfidenzintervall, d.h. das nominale Niveau wird genau eingehalten.

Eine Übertragung des beschriebenen Vorgehens zur Berechnung eines nichtzentralen Konfidenzintervalls auf den Fall heterogener Varianzen ist direkt möglich. Im Modell des Welch-Tests folgt die Teststatistik $t_{emp}^{Welch} = (\bar{X}_1 - \bar{X}_2) / \sqrt{S_1^2/n_1 + S_2^2/n_2}$ einer nichtzentralen t -Verteilung mit $\hat{v} = [q^2/(n_1 - 1) + (1 - q)^2/(n_2 - 1)]^{-1}$ Freiheitsgraden, wobei $q = (S_1^2/n_1) / [(S_1^2/n_1) + (S_2^2/n_2)]$. Auf Basis der Konfidenzgrenzen des Nichtzentralitätsparameters δ'_u und δ'_o , die mit den Statistiken des Welch-Tests analog zur oben beschriebenen Vorgehensweise berechnet werden, kann nun ein Intervall für d' konstruiert werden:

$$d'_u = \delta'_u \cdot \sqrt{\frac{2 \cdot (n_1 \cdot S_1^2 + n_2 \cdot S_2^2)}{n_1 \cdot n_2 \cdot (S_1^2 + S_2^2)}}, \quad d'_o = \delta'_o \cdot \sqrt{\frac{2 \cdot (n_1 \cdot S_1^2 + n_2 \cdot S_2^2)}{n_1 \cdot n_2 \cdot (S_1^2 + S_2^2)}}.$$

Dieses Intervall ist aufgrund der Varianzschätzungen, die in die Bestimmung der Freiheitsgrade sowie die Umrechnung in die Einheit der Effektgröße eingehen, nur ein approximatives, also kein exaktes Intervall mehr. Für ein solches müßten die Varianzen bekannt sein.

Die nichtzentrale Intervallschätzung für d^{S_k} kann ebenfalls nur approximativ vorgenommen werden. Zudem geht in die Beziehung zum Nichtzentralitätsparameter im Falle heterogener Varianzen auch die (geschätzte) Varianz der jeweils anderen, nicht zur Standardisierung herangezogenen Gruppe k' ein (vgl. Algina, Keselman & Penfield, 2006a):

$$d_u^{S_k} = \delta_u^{S_k} \cdot \sqrt{\frac{1}{n_k} + \frac{S_{k'}^2}{n_{k'} S_k^2}}, \quad d_o^{S_k} = \delta_o^{S_k} \cdot \sqrt{\frac{1}{n_k} + \frac{S_{k'}^2}{n_{k'} S_k^2}},$$

wobei $t_{emp}^{S_k} = (\bar{X}_1 - \bar{X}_2) / [S_k \sqrt{1/n_k + S_{k'}^2 / (n_{k'} S_k^2)}]$ und $v = n_k - 1$ sind. Die jeweils analog konstruierten nichtzentralen Intervalle für robustifizierte Effektgrößen werden hier nicht näher dargestellt; die in den Simulationen verwendeten Formeln finden sich in Tabelle 6 (siehe S. 30).

2.1.2.3 Bootstrap-Schätzer

In der vorliegenden Arbeit werden drei verbreitete Bootstrap-Ansätze zur Konstruktion von Konfidenzintervallen verfolgt, das Perzentil-, das BCA- und das Perzentil- t -Verfahren. Das grundsätzliche Vorgehen des Bootstrap-Verfahrens sowie die im folgenden benutzte Notation wurden bereits bei der Beschreibung der Bootstrap-Biaskorrektur (siehe Kapitel 2.1.1.5) eingeführt.

2.1.2.3.1 Perzentil-Verfahren

Die Konfidenzgrenzen werden direkt über Quantile der Bootstrap-Verteilung \hat{F}_n^* berechnet. Als untere $\hat{\theta}_u$ und obere Grenzen $\hat{\theta}_o$ des symmetrischen Konfidenzintervalls zum Niveau $1 - \alpha$ für den Parameter θ werden die Quantile $\hat{F}_n^{*-1}(\alpha/2)$ und $\hat{F}_n^{*-1}(1 - \alpha/2)$ gewählt. Voraussetzung für die Tauglichkeit des Perzentil-Verfahrens ist die Existenz einer vom Parameter unabhängigen monotonen Transformation $g(\cdot)$ mit folgender Eigenschaft (Efron & Tibshirani, 1993, S. 173; Shao & Tu, 1995, S. 132):

$$[g(\hat{\theta}) - g(\theta)] \sim N(0,1).$$

Der Vorteil des Bootstraps ist nun, daß die Beziehung zur Standardnormalverteilung nicht bekannt sein muß, sondern die Annahme ihrer Existenz ausreicht, da die Bootstrap-Technik diese Transformation quasi automatisch durchführt.

2.1.2.3.2 BCA-Verfahren

Eine Verallgemeinerung der Perzentilmethode stellt das BCA-Verfahren dar, wobei die Abkürzung für „bias corrected and accelerated“ steht. Die weniger strikte Annahme dieses Verfahrens sieht neben einer vom Parameter unabhängigen monotonen Transformation $g(\cdot)$ die Existenz zweier Konstanten vor, für die gilt (Efron & Tibshirani, 1993, S. 185):

$$\frac{g(\hat{\theta}) - g(\theta)}{1 + a_{g(\theta)}} + z_0 \sim N(0,1).$$

Hierbei bezeichnen z_0 die Biaskonstante und a die Akzelerationskonstante, mit welcher die vom Parameter abhängige Änderungsrate der Varianz des Schätzers berücksichtigt wird. Die Gewinnung der Bootstrap-Verteilung erfolgt wie beim Perzentil-Verfahren, jedoch als untere Grenze des BCA-Intervalls zum Niveau $1 - \alpha$ wird $\hat{F}_n^{*-1}(\alpha_1)$ und als obere Grenze wird $\hat{F}_n^{*-1}(\alpha_2)$ gewählt, wobei

$$\alpha_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z_{\alpha/2}}{1 - \hat{a}(\hat{z}_0 + z_{\alpha/2})}\right) \quad \text{und} \quad \alpha_2 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z_{1-\alpha/2}}{1 - \hat{a}(\hat{z}_0 + z_{1-\alpha/2})}\right).$$

Für $\hat{a} = 0$ und $\hat{z}_0 = 0$ liefert die BCA-Methode also dieselbe Intervallschätzung wie das Perzentil-Verfahren. Die Schätzung der Bias-Korrektur erfolgt über den Anteil der Bootstrap-Schätzungen, die kleiner sind als die Schätzung aus der konkreten Stichprobe:

$$\hat{z}_0 = \Phi^{-1} \left(\frac{\#(\hat{\theta}_b^* < \hat{\theta})}{B} \right),$$

wobei Φ^{-1} die Inverse der Standardnormalverteilungsfunktion ist und $\#$ für die *Anzahl der Fälle* steht. Wenn der Median der Bootstrap-Verteilung gleich dem Schätzwert der gegebenen Stichprobe ist, nimmt \hat{z}_0 den Wert Null an. Da die Bootstrap-Verteilung bei kleinen Stichproben oder diskreten Daten nicht stetig ist, wurde hier z_0 wie folgt geschätzt:

$$\hat{z}_0 = \Phi^{-1} \left(\frac{\#(\hat{\theta}_b^* < \hat{\theta}) + 0.5\#(\hat{\theta}_b^* = \hat{\theta})}{B} \right).$$

Für die Akzelerationskonstante bieten sich verschiedene Berechnungsmöglichkeiten an, wobei hier die auf dem Jackknife beruhende Variante vorgestellt wird. Das Jackknife-Verfahren ist ein Vorläufer des Bootstraps und wurde von Quenouille (1949) zunächst als Methode zur Bias-Schätzung eingeführt und von Tukey (1958) zu einem universellen Verfahren zur Schätzung von Standardfehlern und Vertrauensintervallen weiterentwickelt. Beim Jackknife werden aus einer gegebenen Stichprobe N Resamples der Größe $N - 1$ ohne Zurücklegen gezogen. Die Jackknife-Replikation eines Schätzers $\hat{\theta}_{(-i)}$ errechnet sich also aus den Werten der Ausgangsstichprobe, von denen der i -te Wert ausgeschlossen wurde. Der Schätzer der Akzelerationskonstante auf Basis der Jackknife-Replikationen lautet nun

$$\hat{a} = \frac{\sum_{i=1}^N (\tilde{\theta} - \theta_{(-i)})^3}{6 \cdot \left(\sum_{i=1}^N (\tilde{\theta} - \theta_{(-i)})^2 \right)^{3/2}}, \text{ wobei } \tilde{\theta} = \sum_{i=1}^N \hat{\theta}_{(-i)} / N.$$

Die Akzeleration des Standardfehlers wird aus der Schiefe der Jackknife-Verteilung geschätzt (Efron, 1987).

2.1.2.3.3 Perzentil- t -Verfahren

Das Perzentil- t -Verfahren kann als Generalisierung der Student- t -Methode betrachtet werden. Hierbei wird in jeder Bootstrap-Stichprobe der studentisierte Schätzer

$$T_n^* = \frac{\hat{\theta}^* - \hat{\theta}}{\sqrt{\text{Var}^*(\hat{\theta})}}$$

berechnet, wobei $\hat{\theta}^*$ der Schätzer des Parameters θ in der jeweiligen Bootstrap-Stichprobe, $\hat{\theta}$ der Schätzer für θ in der konkreten Stichprobe und $\text{Var}^*(\hat{\theta})$ der Schätzer der Varianz von $\hat{\theta}$ in der Bootstrap-Stichprobe sind. Alternativ zur Verwendung der zentralen t -Verteilung wird nun die auf

empirischem Weg gewonnene Verteilung zur Konstruktion eines Konfidenzintervalls herangezogen:

$$\hat{\theta}_u = \hat{\theta} - \hat{c}_{1-\alpha/2} \sqrt{\text{Var}(\hat{\theta})}, \quad \hat{\theta}_o = \hat{\theta} - \hat{c}_{\alpha/2} \sqrt{\text{Var}(\hat{\theta})},$$

wobei $\hat{c}_{\alpha/2}$ und $\hat{c}_{1-\alpha/2}$ Quantile der studentisierten Bootstrap-Verteilung sind.

2.1.3 Fragestellung und ausgewählte Statistiken

In der psychologischen und medizinischen Literatur finden sich zahlreiche Arbeiten zu Intervallschätzern für d , in welchen jedoch in aller Regel nur approximative zentrale Schätzer für den Fall homogener Varianzen behandelt werden. Einen aktuellen Überblick geben z.B. die Arbeiten von Wu, Jiang und Wei (2006) sowie Viechtbauer (2007). Den in der bisherigen Darstellung aufgeworfenen Fragen zur Robustheit, speziell der Unempfindlichkeit gegenüber nichtnormalen Verteilungen sowie der Varianzheterogenität, wird hingegen selten nachgegangen.

Kelley (2005) vergleicht drei Intervallschätzer für d , das nichtzentrale, das Perzentil- und das BCA-Bootstrap-Verfahren unter varianzhomogenen stetigen Verteilungen mit Schiefewerten (γ_1) zwischen 0 und 1.75 und Wölbungsparametern⁵ (γ_2) zwischen -0.75 und 3.75 . Ergebnisse dieser Studie sind, daß das nichtzentrale Verfahren unter nichtnormalen Modellen nicht robust ist sowie daß das BCA-Verfahren besser überdeckt als das Perzentil-Bootstrap und in der Regel das nominale Niveau einhält.

Die Empfehlung des BCA-Verfahrens wird jedoch von Algina, Keselman und Penfield (2006b) aus verschiedenen Gründen kritisiert. Zum einen schneidet dieses Verfahren bei größeren Parameterwerten selbst unter den von Kelley simulierten Modellen nicht immer robust ab; ein Ergebnis, welches Kelley nicht erhalten konnte, da die unabhängigen Variablen N und d in seiner Studie nicht vollständig kombiniert waren. Zum anderen aber halten die Autoren grundsätzlich die Wahl einer nicht robusten Statistik wie d für falsch und erweitern aus dieser Perspektive nun das Untersuchungsdesign um Verteilungsmodelle mit größeren Werten von Schiefe ($0 \leq \gamma_1 \leq 4.9$) und Wölbung ($0 \leq \gamma_2 \leq 4674$). Ihre Argumentation hierfür lautet, daß ein Verfahren, welches sich in Simulationen unter extremen Bedingungen als robust erweist, auch für die Datensituationen der Praxis geeignet ist. Bei diesen Verteilungsmodellen nun unterschreitet auch das von Kelley empfohlene BCA-Verfahren in zahlreichen Bedingungen die nominale Überdeckungswahrscheinlichkeit sehr deutlich und kann für diese nicht mehr als robust beurteilt werden. Als Alternative zu d empfehlen die Autoren daher die robustifizierte Größe d_R^{Algina} und zeigen anhand von Simulationsergebnissen für nichtzentrale sowie Perzentil-Bootstrap-Intervalle, daß mit letzteren eine robuste Schätzung möglich ist. Auf BCA-Bootstrap-Intervalle oder andere Verfahren für d_R^{Algina} gehen die Autoren nicht ein. Diese Ergebnisse zur besseren Überdeckung von d_R^{Algina} anhand des Perzentil-Bootstraps im Vergleich zum nichtzentralen Intervall werden auch durch eine frühere Arbeit der Autoren gestützt (Algina, Keselman & Penfield, 2005a). In dieser werden aber zudem auch die Intervalllängen (im folgenden synonym für Intervallbreite) der beiden Schätzer miteinander verglichen, wobei sich zeigt, daß das nichtzentrale Intervall durchgängig kürzer ist als das Bootstrap-Intervall. Aus diesen beiden und bisher einzigen Studien zu Intervallschätzern von d_R^{Algina} läßt sich somit unmittelbar eine

⁵ Als Wölbungsparameter (Kurtosis, Exzeß) wird im folgenden $\gamma_2 = \mu_4(X)/\sigma^4(X) - 3$ betrachtet, so daß die Wölbung der Normalverteilung gleich Null ist.

Fragestellung für die vorliegende Arbeit ableiten: Wie schneidet das BCA-Verfahren für die robustifizierte Effektgröße im Hinblick auf Überdeckungswahrscheinlichkeit und Intervalllänge ab?

Zur Frage, wie gut die Überdeckung von Cohens d' unter heterogenen Varianzen gelingt, wurden bisher nur wenige Studien durchgeführt (Hess & Kromrey, 2002, 2003, 2004). Da in diesen Arbeiten jedoch zum einen nicht deutlich wird, wie der zu schätzende Parameter genau definiert wurde – obwohl es nicht explizit gesagt wird, kann es sich sinnvollerweise nur um d' handeln – und zum anderen nur Intervalle für d , also auf Basis von d_s , berechnet wurden, fällt das Ergebnis hinsichtlich der Überdeckung dieser Intervalle bei ungleich großen Gruppen erwartungsgemäß schlecht aus. Da keine Gegenüberstellung zu den Intervallen auf Basis von d'_s vorgenommen wurde, wird auf die Ergebnisse hier nicht näher eingegangen. Aus den offenen Fragen dieser Untersuchungen aber kann unmittelbar für die hier angestrebte Studie folgende Fragestellung abgeleitet werden: Wie gut schneiden die Intervallschätzer für d und d' im Vergleich ab? Konkret ist hier interessant, wie groß die Vorteile der Verfahren für d bei Gültigkeit der Voraussetzungen sind. Denn analog zu den einschlägigen Ergebnissen zum Vergleich von t - und Welch-Test (Büning, 1991) läßt sich erwarten, daß die Nachteile der Welch-Verfahren bei homogenen Normalverteilungen gering sind im Vergleich zu den Vorteilen bei Verletzungen der Voraussetzungen insbesondere bei ungleich großen Gruppen.

Die einzige Studie zum Überdeckungsverhalten von Intervallen für d^{S_k} bei Varianzheterogenität wurde ebenfalls von Algina, Keselman und Penfield (2006a) vorgelegt. Das Ergebnismuster dieser Untersuchung stimmt im wesentlichen überein mit den oben berichteten Ergebnissen zu den Konfidenzintervallen für d . Während also die Überdeckung von d^{S_k} mit nichtzentralen Intervallen nur unter der Normalverteilung gelingt und auch das Perzentil-Bootstrap unter tailstarken Verteilungen keine robuste Schätzung garantiert, führt allein das Perzentil-Bootstrap in Verbindung mit robustifizierten Schätzern unter allen simulierten Bedingungen zu einer robusten Überdeckung des Parameters $d_R^{S_k}$. Da auch hier lediglich das Perzentil-Bootstrap als alternatives Verfahren zur Konstruktion der Intervalle eingesetzt wurde, schließt sich für die Planung der folgenden Studie wiederum die Frage nach der Tauglichkeit anderer Bootstrap-Schätzer wie des BCA- und des Perzentil- t -Verfahrens für diese Parameter an. Eine Studie zu den Eigenschaften von Intervallschätzern der robustifizierten Effektgröße d'_R bei heterogenen Varianzen wurde bisher noch nicht vorgelegt.

2.1.3.1 Punktschätzer

Als zu schätzende Parameter (vgl. Tabelle 4) wurden die beiden d -Statistiken von Cohen sowie die an nur einer Streuung relativierte Effektgröße von Glass ausgewählt. Neben der klassischen Version mit KQ-Statistiken wurden zudem robustifizierte Varianten mit einem Trimm- und Winsorisierungsanteil von 20% auf jeder Verteilungsseite in die Untersuchung aufgenommen.

Tabelle 4. Parameter und Punktschätzer der Effektgröße bei zwei unabhängigen Gruppen¹

Parameter	KQ		Parameter ²	(20%-)getrimmt	
		Schätzer			Schätzer ³
$d = \frac{\mu_1 - \mu_2}{\sigma}$	$d_s = \frac{\bar{X}_1 - \bar{X}_2}{S_p}$		$d_R = \frac{\mu_{t1} - \mu_{t2}}{\sigma_w}$	$d_{s,R} = \frac{\bar{X}_{t1} - \bar{X}_{t2}}{S_{wp}}$	
	$\hat{d} = c(n_1 + n_2 - 2)d_s$			$\hat{d}_R = c(h_1 + h_2 - 2)d_{s,R}$	
	$d_s^{Boot} = 2d_s - \frac{1}{B} \sum d_{s,b}^*$			$d_{s,R}^{Boot} = 2d_{s,R} - \frac{1}{B} \sum d_{s,R,b}^*$	
	$\hat{d}^{Boot} = 2\hat{d} - \frac{1}{B} \sum \hat{d}_b^*$			$\hat{d}_R^{Boot} = 2\hat{d}_R - \frac{1}{B} \sum \hat{d}_{R,b}^*$	
$d' = \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}}$	$d'_s = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2 + S_2^2}{2}}}$		$d'_R = \frac{\mu_{t1} - \mu_{t2}}{\sqrt{\frac{\sigma_{w1}^2 + \sigma_{w2}^2}{2}}}$	$d'_{s,R} = \frac{\bar{X}_{t1} - \bar{X}_{t2}}{\sqrt{\frac{S_{w1}^2 + S_{w2}^2}{2}}}$	
	$\hat{d}' = c \left[\frac{(n_1 - 1) \cdot (n_2 - 1) \cdot (S_1^2 + S_2^2)^2}{(n_1 - 1) \cdot S_1^4 + (n_2 - 1) \cdot S_2^4} \right] d'_s$			$\hat{d}'_R = c \left[\frac{(h_1 - 1) \cdot (h_2 - 1) \cdot (S_{w1}^2 + S_{w2}^2)^2}{(h_1 - 1) \cdot S_{w1}^4 + (h_2 - 1) \cdot S_{w2}^4} \right] d'_{s,R}$	
	$d_s'^{Boot} = 2d'_s - \frac{1}{B} \sum d_{s,b}'^*$			$d_{s,R}'^{Boot} = 2d'_{s,R} - \frac{1}{B} \sum d_{s,R,b}'^*$	
	$\hat{d}'^{Boot} = 2\hat{d}' - \frac{1}{B} \sum \hat{d}_b'^*$			$\hat{d}'_R^{Boot} = 2\hat{d}'_R - \frac{1}{B} \sum \hat{d}_{R,b}'^*$	
$d^{S_k} = \frac{\mu_1 - \mu_2}{\sigma_k}$	$d_s^{S_k} = \frac{\bar{X}_1 - \bar{X}_2}{S_k}$		$d_R^{S_k} = \frac{\mu_{t1} - \mu_{t2}}{\sigma_{wk}}$	$d_{s,R}^{S_k} = \frac{\bar{X}_{t1} - \bar{X}_{t2}}{S_{wk}}$	
	$\hat{d}^{S_k} = c(n_k - 1)d_s^{S_k}$			$\hat{d}_R^{S_k} = c(h_k - 1)d_{s,R}^{S_k}$	
	$d_s^{S_k,Boot} = 2d_s^{S_k} - \frac{1}{B} \sum d_{s,b}^{S_k,*}$			$d_{s,R}^{S_k,Boot} = 2d_{s,R}^{S_k} - \frac{1}{B} \sum d_{s,R,b}^{S_k,*}$	
	$\hat{d}^{S_k,Boot} = 2\hat{d}^{S_k} - \frac{1}{B} \sum \hat{d}_b^{S_k,*}$			$\hat{d}_R^{S_k,Boot} = 2\hat{d}_R^{S_k} - \frac{1}{B} \sum \hat{d}_{R,b}^{S_k,*}$	

¹ Die t -basierte Biaskorrektur lautet: $c(m) = \Gamma(m/2) / [(m/2)^{-0.5} \Gamma((m-1)/2)]$.

² Auf die Multiplikation mit der 20%-winsorisierten Streuung der Normalverteilung wird hier verzichtet, da in der Auswertung die Vergleichbarkeit zwischen KQ- und robustifizierten Maßen je Verteilungsmodell eigens herzustellen ist (vgl. Kapitel 2.1.5.2).

³ $h_k = n_k - 2\gamma n_k$.

2.1.3.2 Intervallschätzer

Als Verfahren zur Intervallschätzung wurden für jeden Parameter ein nichtzentrales Intervall auf Basis der t -Verteilung sowie das Perzentil-, das BCA- und das Perzentil- t -Bootstrap ausgewählt. Da die Bootstrap-Verfahren jeweils sowohl mit dem nichtkorrigierten als auch mit dem auf Basis der t -Verteilung korrigierten Schätzer simuliert wurden, ergeben sich insgesamt für jeden Parameter

sieben Intervallschätzer. Zunächst sei exemplarisch für d' in Tabelle 5 die auch für alle anderen Intervallschätzer benutzte Notation, wie sie im Ergebnisteil verwendet wird, eingeführt.

Tabelle 5. Notation der ausgewählten Intervallschätzer am Beispiel von d'

Intervallschätzer	Beschreibung des Verfahrens
NC (d'_s)	Nichtzentrales Intervall auf Basis der t -Verteilung
Perc (d'_s)	Perzentil-Bootstrap mit nichtkorrigiertem Punktschätzer
BCA (d'_s)	BCA-Bootstrap mit nichtkorrigiertem Punktschätzer
Perc- t (d'_s)	Perzentil- t -Bootstrap mit nichtkorrigiertem Punktschätzer
Perc (\hat{d}')	Perzentil-Bootstrap mit korrigiertem Punktschätzer
BCA (\hat{d}')	BCA-Bootstrap mit korrigiertem Punktschätzer
Perc- t (\hat{d}')	Perzentil- t -Bootstrap mit korrigiertem Punktschätzer

Alle Bootstrap-Verfahren wurden mit $B = 1,000$ Wiederholungsstichproben durchgeführt. Während für die Berechnung des Perzentil- sowie des BCA-Verfahrens nur der Punktschätzer auszuwählen ist, muß für das Perzentil- t -Bootstrap zudem ein Varianzschätzer für die Studentisierung gefunden werden.

Die Konstruktion von Varianzschätzern für standardisierte Lagedifferenzen kann direkt über die Beziehung zur jeweiligen t -Statistik erfolgen (Hedges & Olkin, 1985). Da beispielsweise für den klassischen t -Test gilt, daß $d_s \cdot \sqrt{(n_1 \cdot n_2)/(n_1 + n_2)}$ mit $v = n_1 + n_2 - 2$ Freiheitsgraden und Nichtzentralitätsparameter $\delta = d \cdot \sqrt{(n_1 \cdot n_2)/(n_1 + n_2)}$ t -verteilt ist, kann die Varianz von d_s über $\sigma_{d_s}^2 = [(n_1 + n_2)/(n_1 \cdot n_2)] \cdot \sigma_{t(v,\delta)}^2$ berechnet werden, wobei $\sigma_{t(v,\delta)}^2$ die Varianz der nichtzentralen t -Verteilung ist (Johnson & Welch, 1939):

$$\begin{aligned}\sigma_{t(v,\delta)}^2 &= \frac{v}{v-2} (1 + \delta^2) - \frac{v}{2} \delta^2 \left[\frac{\Gamma((v-1)/2)}{\Gamma(v/2)} \right]^2 \\ &= \frac{v}{v-2} (1 + \delta^2) - \frac{\delta^2}{[c(v)]^2}.\end{aligned}$$

In Anlehnung an die statistische Literatur zur Metaanalyse wird im folgenden die Varianzdarstellung in Abhängigkeit von der Effektgröße gewählt

$$\sigma_{\hat{\theta}}^2 = \frac{1}{f} \cdot \frac{v}{v-2} \cdot (1 + f \cdot \theta^2) - \frac{\theta^2}{[c(v)]^2},$$

wobei f die Beziehung zwischen Effektgröße und Nichtzentralitätsparameter angibt: $f = \delta^2/\theta^2$. Sofern die Effektgröße mit einem Faktor $c(m)$ biaskorrigiert wird, ergibt sich z.B. $\sigma_{\hat{d}}^2 = [c(m)]^2 \cdot \sigma_{d_s}^2$, wobei $\hat{d} = c(m) \cdot d_s$. Analog lassen sich auch die Varianzen der Effektgrößen, die mit dem Welch-Test und den getrimmten t -Tests (Yuen, 1974; Yuen & Dixon, 1973) assoziiert sind, bestimmen. Zur Schätzung der Varianz wird jeweils der unbekannte Effektgrößenparameter durch den Schätzer der Effektgröße $\hat{\theta}$ ersetzt, so daß der Varianzschätzer einer Effektgröße $\hat{\theta}$, die eine Funktion einer t -verteilten Statistik ist, allgemein wie folgt lautet:

$$\hat{\sigma}_0^2 = \frac{1}{f} \cdot \frac{v}{v-2} \cdot (1 + f \cdot \hat{\theta}^2) - \frac{\hat{\theta}^2}{[c(v)]^2}.$$

Tabelle 6 faßt für die hier betrachteten Varianten des t -Tests die Beziehungen zwischen den Effektgrößen und der jeweiligen t -Statistik zusammen. Die Formeln für f und v ermöglichen in Verbindung mit $\hat{\sigma}_0^2$ also die Berechnung der Standardfehler von Effektgrößen im Rahmen der verschiedenen t -Tests und dienen zudem in Verbindung mit dem in Kapitel 2.1.2.2 beschriebenen Vorgehen zur Bestimmung der nichtzentralen Intervalle.

Tabelle 6. Beziehungen zwischen Effektgrößen und t -Statistik

	$\hat{\theta}$	f	v
t -Test	$d_s; \hat{d}$	$\frac{n_1 + n_2}{n_1 \cdot n_2}$	$n_1 + n_2 - 2$
Welch-Test	$d'_s; \hat{d}'$	$\frac{2 \cdot (n_2 \cdot S_1^2 + n_1 \cdot S_2^2)}{n_1 \cdot n_2 \cdot (S_1^2 + S_2^2)}$	$\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right)^2 \left/ \left(\frac{S_1^4}{n_1^2(n_1-1)} + \frac{S_2^4}{n_2^2(n_2-1)} \right) \right.$
Algina ¹	$d_s^{S_k}; \hat{d}^{S_k}$	$\frac{1}{n_k} + \frac{S_k^2}{n_k \cdot S_k^2}$	$n_k - 1$
getrimmter t -Test	$d_{s,R}; \hat{d}_R$	$\frac{(h_1 + h_2)(n_1 + n_2 - 2)}{h_1 h_2 (h_1 + h_2 - 2)}$	$h_1 + h_2 - 2$
getrimmter Welch-Test ²	$d'_{s,R}; \hat{d}'_R$	$\left(\frac{SSD_{w1}}{h_1^2 - h_1} + \frac{SSD_{w2}}{h_2^2 - h_2} \right) \left/ \left(\frac{SSD_{w1}}{2n_1 - 2} + \frac{SSD_{w2}}{2n_2 - 2} \right) \right.$	$\left(\frac{S_{w1}^2}{h_1} + \frac{S_{w2}^2}{h_2} \right)^2 \left/ \left(\frac{S_{w1}^4}{h_1^2(h_1-1)} + \frac{S_{w2}^4}{h_2^2(h_2-1)} \right) \right.$
Algina robustifiziert	$d_{s,R}^{S_k}; \hat{d}_R^{S_k}$	$\frac{n_k - 1}{h_k - 1} \left(\frac{1}{h_k} + \frac{SSD_{wk'} / (h_k - 1)}{h_k \cdot SSD_{wk'} / (h_k - 1)} \right)$	$h_k - 1$

¹ Algina, Keselman & Penfield (2006a).

² SSD = sum of squares.

2.1.4 Simulationsmethode

Im folgenden wird die Auswahl und Erzeugung der simulierten Verteilungen beschrieben, die Anzahl der Simulationsdurchgänge je Untersuchungsbedingung diskutiert sowie das Design der Studie vorgestellt.

2.1.4.1 Verwendete Verteilungen

Zur Untersuchung der Robustheit von Statistiken bieten sich mehrere Verteilungsfamilien – sogenannte Supermodelle (Büning, 1991) – an, die jeweils unterschiedliche Klassen von Verteilungen umfassen. Gemeinsam ist diesen Modellen, daß sie zur Erzeugung stetiger Verteilungen dienen und alle die Normalverteilung einschließen. Büning (S. 16 f.) nennt vier solcher Supermodelle: (1) die Pearson-Verteilungen, (2) die Verteilungen von Box und Tiao, (3) die Ramberg-Schmeiser-Tukey-Verteilungen (R-S-T) sowie (4) die kontaminierte Normalverteilung (KNV). Letztere ist ein Spezialfall des gross-error-Modells, bei welchem zwei beliebige Verteilungen gemischt werden. Dieser Liste lassen sich weitere Familien hinzufügen, zu nennen sind vor allem die g -und- h -Verteilungen

(Hoaglin, 1985) und die Verteilungen von Fleishman (1978), die jeweils auch in der psychologischen Literatur häufiger Verwendung finden. Die Unterschiede zwischen diesen Supermodellen bestehen in der Größe und Form des Raumes, in welchem sich die realisierbaren Verteilungsmomente bewegen. Neben den Modellen für stetige Verteilungen, stehen für Simulationen sämtliche (theoretischen) diskreten Verteilungen zur Verfügung, wobei diese vergleichsweise selten zur Anwendung kommen.

Die Auswahl der Verteilungen sollte nun einerseits einen weiten Bereich von statistischen Eigenschaften abdecken. Zu beachten ist aber, daß die verfügbaren Klassifikationsmaße wie Schiefe, Wölbung, Peakedness und Tailstärke die Vergleichbarkeit von Simulationsergebnissen unter verschiedenen Supermodellen nicht garantieren. So haben bereits Lee und Gurland (1977) zeigen können, daß das Testniveau des t -Tests unter Zugrundelegung verschiedener Verteilungen, die in den ersten vier Momenten übereinstimmen, erheblich schwanken kann. Die simulierte Robustheit einer Statistik ist also grundsätzlich modellabhängig. Andererseits sollten die simulierten Verteilungen auch den Gegebenheiten empirischer Verteilungen nahekommen, eine Forderung, die im Hinblick auf die praktische Bedeutung von Simulationen vielleicht sogar wichtiger ist. Für den sozialwissenschaftlichen Bereich hat vor allem Micceri (1989, 1990) eine Untersuchung der Charakteristika von 440 großen empirischen Verteilungen vorgelegt. Wesentliche Ergebnisse dieser Studie sind:

- Die Normalverteilungsannahme von Rohwerten ist nahezu nie haltbar.
- Empirische Daten sind in der Regel asymmetrisch, wobei die Schiefe selten diejenige der Exponentialverteilung überschreitet.
- Die Tailstärken liegen zwischen denen der Gleichverteilung und der Doppelexponentialverteilung.
- Empirische Verteilungen sind vielfach diskret, multimodal und klumpig („lumpy“) oder weisen gehäuft einzelne Werte auf.

Da sich diese Charakteristika nicht mit den Eigenschaften decken, die typischerweise in Simulationsstudien erzeugt werden, kommt Micceri insgesamt zu dem Urteil, daß eine Neubewertung der Robustheitsliteratur erforderlich ist. Eine vergleichbare Kritik hatte zuvor bereits Stigler (1977) an der sogenannten Princeton-Studie (Andrews, Bickel, Hampel, Huber, Rogers & Tukey, 1972) geübt, in welcher die Welt durch „Cauchy-colored glasses“ (S. 1070) betrachtet worden sei. Die weitgehende Beschränkung auf symmetrische Verteilungen und die Fokussierung auf Verteilungen mit extrem starken Tails läßt nach Stigler nicht auf die praktische Relevanz der in dieser Studie untersuchten Lagestatistiken schließen. Zwar kommen neuere Studien dieser Kritik im wesentlichen nach, d.h. es werden auch asymmetrische und weniger tailstarke Verteilungen simuliert, jedoch läßt sich bei Simulationen zur robusten Statistik insbesondere in psychologischen Zeitschriften weiterhin die Tendenz entdecken, daß Verteilungen ausgewählt werden, deren Tailstärke deutlich oberhalb derjenigen der Normalverteilung liegt (z.B. Luh & Guo, 1999; Wilcox, Keselman & Kowalchuk, 1998). Dies ist jedoch nicht mit der auch von diesen Autoren zitierten Arbeit von Micceri in Einklang zu bringen, der immerhin bei 18% der 440 untersuchten empirischen Verteilungen kürzere Tails als bei der Normalverteilung festgestellt hat. Zudem liegt der Verdacht nahe, daß diese Verteilungen bewußt ausgeblendet werden, da die Varianz robuster Statistiken wie z.B. des getrimmten Mittels bei Verteilungen mit schwachen Tails relativ groß ist und im Vergleich zum herkömmlichen Mittelwert mit schlechteren Schätz- und Testeigenschaften einhergeht.

Mit dem Ziel, den genannten Forderungen weitgehend nachzukommen und zudem auch eine Vergleichbarkeit mit aktuellen Untersuchungen zur Effektgröße zu erreichen, wurden die in Tabelle

7 beschriebenen stetigen Verteilungen sowie als diskretes Modell die negative hypergeometrische Verteilung (Hyper) ausgewählt. Zur Bezeichnung der Verteilungen werden im folgenden die in eckigen Klammern angegebenen Kürzel verwendet. Die Momente der $g-h$ -Verteilungen wurden nach den Formeln von Hoaglin (1985, S. 502 f.) berechnet. Die robusten Statistiken wurden per Simulation auf Basis von 10,000 Stichproben der Größe $N = 20,000$ bestimmt.

Tabelle 7. Momente und robustifizierte Parameter der stetigen Verteilungen⁶

		μ	σ	γ_1	γ_2	$\mu_{t, 20\%}$	$\sigma_{w, 20\%}$
[Normal]	Standardnormal	0	1	0	0	0	0.642
[Gleich]	Gleich (-1.732, 1.732)	0	1	0	-1.2	0	0.805
[$g-h-1$]	$g-h$ (0, 0.225)	0	1.566	0	155	0	0.688
[$g-h-2$]	$g-h$ (0.76, -0.098)	0.378	1.252	2	6	0.082	0.669
[$g-h-3$]	$g-h$ (0.225, 0.225)	0.168	1.687	4.9	4674	0.026	0.694
[$g-h-4$]	$g-h$ (0.834, -0.362)	0.299	0.883	1	0	0.091	0.616

Die stetigen Modelle umfassen also drei symmetrische Verteilungen mit kurzen (Gleich), normalen (Normal) und starken ($g-h-1$) Enden sowie drei schiefe Verteilungen, deren Tails von normal ($g-h-4$) bis sehr stark ($g-h-3$) variieren. Abbildung 5 stellt die Verteilungsdichten dar, wobei μ jeweils auf Null gesetzt wurde.

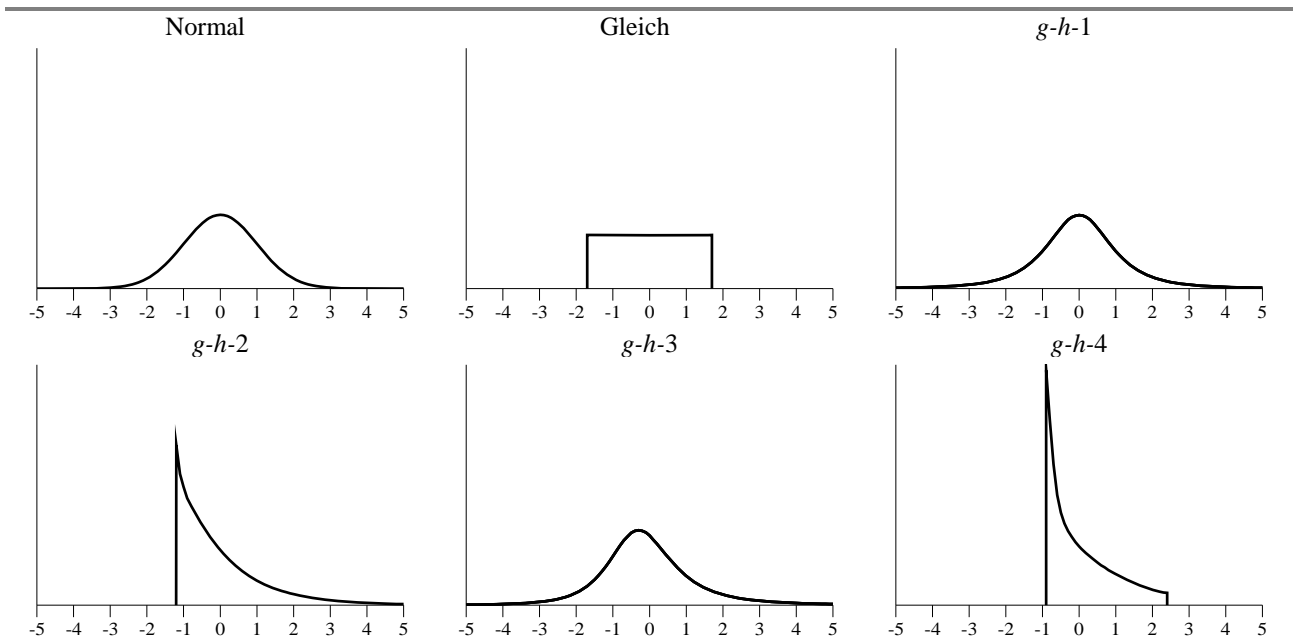


Abbildung 5. Funktionsgraphen der stetigen Verteilungen

Als Modell diskreter Daten wurde die negative hypergeometrische Verteilung – auch bekannt als Beta-Binomialverteilung – herangezogen, wie sie von Lord und Novick (1968, S. 515 ff.) beschrieben wird. Ausgehend von einem binomialen Fehlermodell sowie einer Betaverteilung der wahren

⁶ Da die Wölbung kein geeignetes Maß zur Beschreibung der Tailstärke und Peakedness einer Verteilung ist, die Verwendung der hierfür entwickelten Maße (Bünig, 1991) jedoch wenig verbreitet ist, sind im Anhang B (Tabelle B-1) die Eigenschaften der simulierten Verteilungen anhand von Häufigkeitsverteilungen dargestellt.

Werte beschreibt die negative hypergeometrische Verteilung die Wahrscheinlichkeit eines Summenscores in Abhängigkeit von der Reliabilität einer Skala mit n Items. Als Anzahl der Items wurde $n = 10$ gewählt, der Wertebereich aller negativen hypergeometrischen Verteilungen wurde zudem auf Null bis Zehn festgesetzt. Anders als bei den stetigen Verteilungen wurden die Effektgrößenparameter nicht über lineare Transformationen der zugrundeliegenden Verteilungen gesteuert, sondern über eine geeignete Wahl der Verteilungsparameter, welche die Form der Wahrscheinlichkeitsverteilung festlegen. Aufgrund dieser praxisnahen Festsetzung des Wertebereiches wurden im diskreten Fall also in der Regel, d.h. außer im varianzhomogenen Fall mit Effektgröße gleich Null, zwei Verteilungen mit unterschiedlicher Form gegenübergestellt. Abbildung 6 zeigt exemplarisch für vier Simulationsbedingungen die jeweils den beiden Gruppen zugrundeliegenden negativen hypergeometrischen Wahrscheinlichkeitsverteilungen.

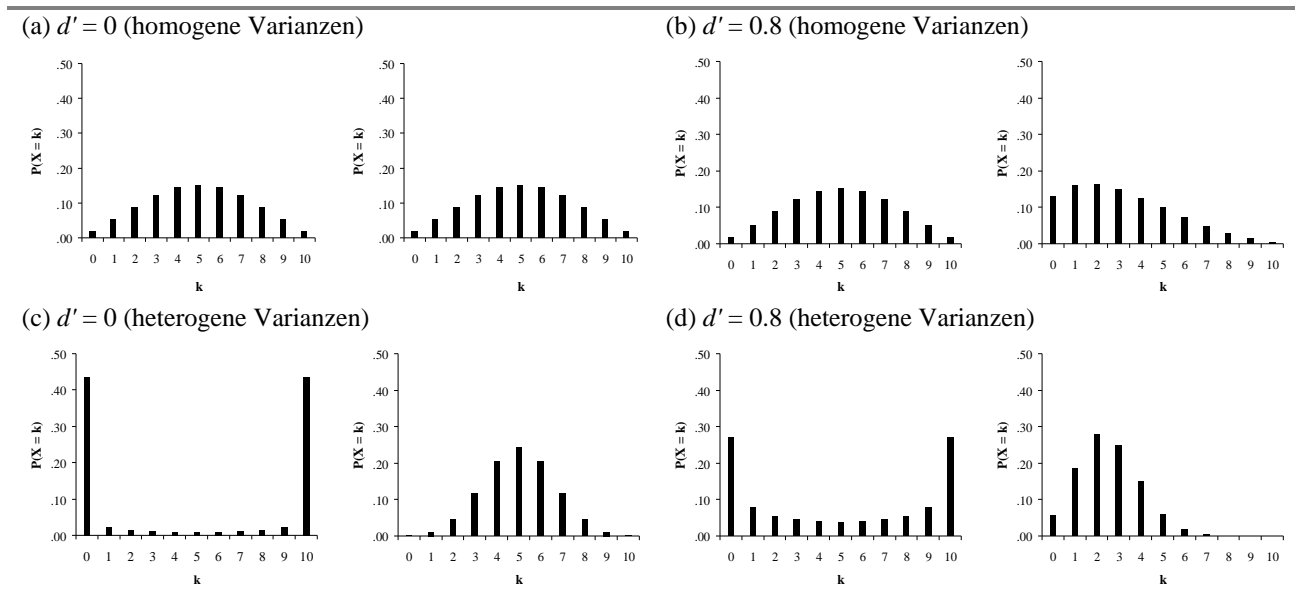


Abbildung 6. Negative hypergeometrische Verteilungen von vier Simulationsbedingungen

2.1.4.2 Erzeugung der Pseudozufallszahlen

Die Simulationen wurden in der Programmumgebung von SPSS 15.0.1.1 realisiert. Als Pseudozufallszahlengenerator ist darin der Mersenne-Twister (Matsumoto & Nishimura, 1998) integriert, der für stochastische Simulationen neben den Verfahren nach L'Ecuyer (1998) zu den empfehlenswerten Erzeugern gleichverteilter Zahlen zählt (Wichmann & Hill, 2006). Zur Erzeugung normalverteilter Zahlen wurde die Methode nach Box und Muller (1958) verwendet, mit der sich aus zwei (0, 1)-gleichverteilten Zufallszahlen R_1 und R_2 zwei stochastisch unabhängige standardnormalverteilte Zahlen Z_1 und Z_2 gewinnen lassen:

$$Z_1 = \sqrt{-2 \ln(1 - R_1)} \cos(2\pi R_2) \quad \text{und} \quad Z_2 = \sqrt{-2 \ln(1 - R_1)} \sin(2\pi R_2).$$

Ausgangspunkt von g - h -Verteilungen (Hoaglin, 1985) ist die Standardnormalverteilung, die über folgende Transformationen mit den Konstanten g hinsichtlich Schiefe und h hinsichtlich Tailstärke manipuliert wird:

$$X = \frac{\exp(gZ) - 1}{g} \exp\left(\frac{hZ^2}{2}\right).$$

Sofern $g = 0$ ist X symmetrisch und wird erzeugt über:

$$X = Z \exp\left(\frac{hZ^2}{2}\right).$$

Wenn $g = h = 0$ ist X standardnormalverteilt. Je größer der Wert für g , desto größer ist die Schiefe von X und je größer h gewählt wird, desto stärker sind die Tails. Die hier gewählten Werte für g und h sowie die zugehörigen Verteilungseigenschaften sind bereits oben in Tabelle 7 aufgeführt.

Die Wahrscheinlichkeitsverteilung der negativen hypergeometrischen Verteilung lautet:

$$h(x) \equiv \frac{b^{[n]}}{(a+b)^{[n]}} \frac{(-n)_x (a)_x}{(-b)_x x!}, \text{ wobei}$$

$$b^{[n]} = b(b-1)\dots(b-n+1), (a)_x = a(a+1)\dots(a+x-1) \text{ und } x = 0, 1, \dots, n.$$

Die negativen hypergeometrischen Zufallszahlen mit $n = 10$ wurden jeweils nach Festlegung der Werte für a und b über die Inversionsmethode, d.h. aus $(0, 1)$ -gleichverteilten Zahlen und der gegebenen inversen Verteilungsfunktion, berechnet. Im Anhang B sind alle in den Simulationen eingesetzten Parameterwerte von a und b aufgeführt. Die ersten beiden Momente dieser Verteilung ergeben sich nach folgenden Formeln, die hier zur Bestimmung der Parameter a und b genutzt werden:

$$\mu = na/D,$$

$$\sigma^2 = \frac{na(D-a)(D+n)}{D^3 + D^2},$$

wobei $D \equiv a + b - n + 1$, $a = (-1 + 1/\alpha_{21})\mu$, $b = -a - 1 + n/\alpha_{21}$ und α_{21} die Kuder-Richardson Formel 21 der Reliabilität (Lord & Novick, 1968, S. 91) bezeichnet.

2.1.4.3 Datenmodelle

Für alle stetigen Verteilungssituationen wird von folgendem Datenmodell ausgegangen, welches Lage- und Skalenunterschiede einschließt: Es seien X_{1i}, \dots, X_{1n_1} und X_{2j}, \dots, X_{2n_2} Zufallsvariablen mit

$$X_{1i} \sim F\left(\frac{x - \mu_1}{\sigma_1}\right) \text{ und } X_{2j} \sim F\left(\frac{x - \mu_2}{\sigma_2}\right),$$

wobei $i = 1, \dots, n_1$, $j = 1, \dots, n_2$ und F stetig mit zugehöriger Dichte f sind. Für die Simulationen bedeutet dies konkret, daß sämtliche Daten innerhalb einer Bedingung aus derselben Basisverteilung mit Erwartungswert Null gezogen werden, danach wird gegebenenfalls die Varianz einer der

beiden Gruppen über einen Faktor verändert und schließlich wird die Lage dieser Gruppe um eine additive Konstante verschoben (Mittelwertsshift), so daß der gewünschte Parameter d' in diesen simulierten Populationsverteilungen realisiert wird.

Bei den diskreten negativen hypergeometrischen Verteilungen, die als Modell eines Testscores fungieren, wird hingegen anders vorgegangen, da die beschriebenen linearen Transformationen den Wertebereich der simulierten Testwertverteilungen verändern würden. Dies erscheint jedoch für die Beurteilung der praktischen Tauglichkeit bei diskreten Daten als unrealistisch. Anders als z.B. bei Algina, Keselman und Penfield (2005a), die auch mit diesem diskreten Modell arbeiten, wurden die Verteilungen hier nicht verschoben, sondern es wurden für jede Situation gesondert Parameter $(a_1; b_1)$ und $(a_2; b_2)$ über Bisektionen bestimmt, für die sich eine gewünschte Effektgröße ergibt (vgl. Anhang B). Das Modell der negativen hypergeometrischen Verteilungen kann wie folgt notiert werden:

$$X_{1i} \sim Nh(n = 10; a_1; b_1) \text{ und } X_{2j} \sim Nh(n = 10; a_2; b_2).$$

Sofern der Effektgrößenparameter ungleich Null oder die Varianzen verschieden sind, werden hier Verteilungen betrachtet, die sich hinsichtlich ihrer Form unterscheiden. Damit sind aber auch bereits die Nachteile dieser Modellierung für die Ergebnisinterpretation angesprochen, denn mit der Manipulation der interessierenden unabhängigen Variablen (Effektgröße und Varianzverhältnis) werden zugleich weitere relevante Verteilungseigenschaften geändert, weshalb das Verhalten der untersuchten Statistiken nicht klar der Veränderung einer Variablen zugeordnet werden kann.

2.1.4.4 Robustheitskriterien und Anzahl der Simulationsdurchgänge

Im Rahmen von Simulationsstudien kommt als Maß der α -Robustheit die relative Abweichung des aktuellen Niveaus α^* vom nominalen Niveau α in Betracht: $r_1(\alpha, \alpha^*) = |\alpha - \alpha^*| / \alpha$. Ein Test oder ein Intervall wird dann als α -robust bezeichnet, wenn $r_1(\alpha, \alpha^*) \leq d_1$. In der Psychologie haben vor allem das „strenge Kriterium“ $d_1 = 0.1$ und das „liberale Kriterium“ $d_1 = 0.5$ von Bradley (1978) Verbreitung gefunden. Die Genauigkeit, mit der in einer Simulation die α -Robustheit nun beurteilt werden kann, hängt von α und der Anzahl der Simulationsdurchgänge N_{MC} ab, wobei die Beurteilung des beobachteten Niveaus $\hat{\alpha}^*$ über Konfidenzintervalle oder auch mit Hilfe von Signifikanztests erfolgen kann.

Teststärkeanalysen werden insbesondere für die Festlegung von N_{MC} in der Studienplanung sowie für die Ableitung kritischer Werte empfohlen. Robey und Barcikowski (1992) kritisieren, daß herkömmliche Simulationsstudien in der Regel keine Begründung für die Wahl von N_{MC} geben und vielfach mit zu wenigen Simulationsdurchgängen durchgeführt werden. Sie weisen aufgrund ihrer Analysen darauf hin, daß optimierte Testungen strenger Robustheitskriterien enorm große Iterationsanzahlen erfordern. Die diesen Teststärkeanalysen und den daraus abgeleiteten Empfehlungen zur Wahl von N_{MC} zugrundeliegenden Hypothesen $H_0: r_1(\alpha, \alpha^*) = 0$ und $H_1: r_1(\alpha, \alpha^*) = d_1$ erscheinen jedoch problematisch (Serlin, 2000). Zum einen ist die exakte Nullhypothese nicht angebracht, da bei der Robustheitsuntersuchung gefragt wird, ob das aktuelle Niveau nahe dem nominalen Niveau ist; die Identität beider Niveaus hingegen wird mit der Definition der Robustheit nicht gefordert und ist in aller Regel auch nicht gegeben. Zum anderen ist die Entscheidung für die Robustheit eines Verfahrens die wissenschaftlich bedeutsamere Aussage gegenüber der Entscheidung für die Nichtrobustheit und erfordert nach Serlin eine strenge Kontrolle der zugeordneten Fehlerwahrscheinlichkeit. Daher sollte mit der Nullhypothese die Nichtrobustheit und mit der

Alternativhypothese die Robustheit angenommen werden. Serlin schlägt zur Testung der Robustheit folgende Bereichsnulhypthesen vor: $H_0^1: \alpha^* \leq \alpha_1$ und $H_0^2: \alpha^* \geq \alpha_2$, wobei $\alpha_1 = \alpha - d_1\alpha$ und $\alpha_2 = \alpha + d_1\alpha$. Für die Testentscheidung sind nun zwei kritische Werte CV_1 und CV_2 festzulegen, für die gilt: $P(CV_1 \leq \hat{\alpha} \leq CV_2 | \alpha^* = \alpha_1) = \omega$ und $P(CV_1 \leq \hat{\alpha} \leq CV_2 | \alpha^* = \alpha_2) = \omega$, wobei ω die Typ-I-Fehlerwahrscheinlichkeit des Tests des Anteilswertes ist. Die Alternativhypothese dieses Tests kann mit $H_1: r_1(\alpha, \alpha^*) = d_1'$ notiert werden, wobei $d_1' < d_1$ und also ein strengeres Kriterium als das in den Nullhypothesen verwendete ist. Hierdurch wird ein weiterer Mangel der Testung nach Robey und Barcikowski beseitigt, daß nämlich mit d_1 das unter Betracht stehende Robustheitskriterium in der Alternativhypothese verwendet wird. Ein Verfahren, für welches $r_1(\alpha, \alpha^*) = d_1$ gilt, ist der Definition nach robust, würde aber im Anteilswerttest nach Robey und Barcikowski mit Wahrscheinlichkeit $1-\beta$ (β ist die Typ-II-Fehlerwahrscheinlichkeit im Signifikanztest nach Robey und Barcikowski) als nicht robust beurteilt werden.

In Anlehnung an Cochran (1952) schlägt Serlin zudem vor, daß das Robustheitskriterium eine Funktion des nominalen Niveaus sein sollte, da eine 50%-Abweichung bei $\alpha = .001$ akzeptabel erscheint, im Bereich von $\alpha = .1$ hingegen nicht. Als Kompromiß zwischen den verschiedenen herkömmlichen Robustheitskriterien gibt er folgende Richtwerte an: $d_{1(\alpha = .01)} = 0.49$, $d_{1(\alpha = .05)} = 0.25$ und $d_{1(\alpha = .1)} = 0.19$ (S. 237). In Tabelle 8 sind nun beispielhaft Teststärken des skizzierten Bereichsnulhypthesentests für das nominale Niveau $\alpha = .05$ bei verschiedenen Null- und Alternativhypothesen aufgeführt, wobei wegen der Diskretheit der Stichprobenverteilung für das Typ-I-Fehlerniveau jeweils $\omega \leq .05$ gilt (die tatsächlichen Testniveaus werden für die Hypothesen der oberen Tabellenhälfte in den beiden letzten Spalten angezeigt). Sofern z.B. Serlins Kriterium $d_1 = 0.25$ zur Ableitung der Nullhypothesen mit $\alpha_1 = .0375$ und $\alpha_2 = .0625$ dient, kann bei den aufgeführten Stichprobengrößen von den hier betrachteten Alternativhypothesen Bradleys Kriterium mit befriedigender Stärke getestet werden. Diese liegt für $N_{MC} = 10,000$ bei zumindest .9339, wobei die Entscheidung für die Robustheit im Sinne des strengen Kriteriums nach Bradley im Rahmen einer Simulation zu fällen wäre, wenn das geschätzte Testniveau eines Verfahrens zwischen .0408 und .0584 läge. Wenn die liberaleren Kriterien nach Cochran oder Serlin mit ähnlicher Stärke getestet werden sollen, muß entweder die Anzahl der Simulationsdurchgänge gesteigert werden oder der Abstand der Nullhypothesen von α muß vergrößert werden (siehe hierzu die untere Hälfte von Tabelle 8).

Tabelle 8. Teststärkeanalysen des Bereichsnulhypthesentests für $\alpha = .05$ nach Serlin (2000)

N_{MC}	Nullhypothesen		Kritische Werte ²		Teststärken verschiedener Alternativhypothesen					
	α_1	α_2	CV_1	CV_2	Bradley: $d_1' = .1$		Cochran: $d_1' = .2$		Serlin ¹ : $d_1' = .25$	
					$P_{(.045)}$	$P_{(.055)}$	$P_{(.04)}$	$P_{(.06)}$	$P_{(.0375)}$	$P_{(.0625)}$
1,000	.0375	.0625	.0490	.0510	.1306	.1348	.0524	.0682	.0249	.0421
2,500	.0375	.0625	.0444	.0544	.5590	.4627	.1422	.1267	.0416	.0491
5,000	.0375	.0625	.0422	.0568	.8387	.7242	.2229	.1783	.0454	.0494
10,000	.0375	.0625	.0408	.0584	.9810	.9339	.3485	.2581	.0450	.0461
1,000	.0300	.0700	.0400	.0560	.7540	.5774	.5167	.3242	.3596	.2180
2,500	.0300	.0700	.0360	.0612	.9887	.9180	.8587	.6202	.6680	.4146
5,000	.0300	.0700	<u>.0342</u>	<u>.0640</u>	.9999	.9971	.9850	.8881	.8985	.6823
10,000	.0300	.0700	.0329	.0657	1.000	1.000	.9999	.9916	.9936	.9095

¹ Die kursiven Werte in der oberen Tabellenhälfte zeigen keine Teststärke, sondern das tatsächliche Testniveau an, da hier $d_1 = d_1' = .25$.

² Die Unterstreichungen zeigen die für die folgenden Studien gewählten kritischen Werte an.

Eine andere Perspektive bietet die Betrachtung der Intervalle von Anteilswerten. In Tabelle 9 sind 95%-Schwankungsintervalle von Anteilswerten in Abhängigkeit vom Wahrscheinlichkeitswert p und der Anzahl der Simulationsdurchgänge N_{MC} aufgeführt.

Tabelle 9. 95%-Schwankungsintervalle von Anteilswerten

N_{MC}	$p = .9300$		$p = .9500$		$p = .9700$	
	UG	OG	UG	OG	UG	OG
1,000	.9130	.9450	.9350	.9630	.9580	.9800
2,500	.9196	.9400	.9408	.9584	.9628	.9764
5,000	.9226	.9370	.9436	.9560	.9650	.9746
10,000	.9249	.9350	.9456	.9542	.9665	.9733

Wie groß sollte die Anzahl der Simulationsdurchgänge nun gewählt werden? Im Gegensatz zu Serlin (S. 235), der wegen der Gefahr übermäßiger Teststärke vor der Wahl zu großer Stichproben in Monte-Carlo-Studien warnt, lautet m.E. die grundsätzliche Antwort hierauf: So groß wie möglich! Denn je größer N_{MC} ist, desto präziser ist die Schätzung und desto geringer kann der Abstand zwischen Null- und Alternativhypothese gewählt werden. Während Simulationen zu klassischen Inferenzverfahren bzw. allgemein zu Verfahren, die über geschlossene Formeln berechnet werden, mit vergleichsweise geringem Rechenaufwand verbunden sind und sich mit großen Replikationsanzahlen ($N_{MC} \geq 10,000$) leicht realisieren lassen, ist die Wahl von N_{MC} bei Studien zu computerintensiven Verfahren wie dem Bootstrap auch heute noch abhängig von der verfügbaren Rechenkapazität. Aufgrund von Voruntersuchungen wurde das N_{MC} der folgenden Untersuchungen einheitlich auf 5,000 festgesetzt. Als nicht zulässige Abweichungen von $\alpha = .05$ wurden $\alpha_1 = .03$ und $\alpha_2 = .07$ festgelegt, so daß für $N_{MC} = 5,000$ gemäß Tabelle 8 als kritische Werte zur Beurteilung der Robustheit .034 und .064 bzw. als kritische Überdeckungswahrscheinlichkeiten .936 und .966 gewählt wurden (vgl. Unterstreichungen in Tabelle 8).

2.1.4.5 Untersuchungsdesign

Aufgrund des erheblichen Rechenaufwandes des Bootstraps wurden die ausgewählten Verfahren nur für zwei Stichprobengrößen simuliert, nämlich $N = 40$ und $N = 120$. Als weitere unabhängige Variablen wurden die Effektgröße d' mit fünf Stufen (0, 0.4, 0.8, 1.2 und 1.6), das Streuungsverhältnis $\sigma_1 : \sigma_2$ mit zwei Stufen (1:1 und 3:1) und das Verhältnis der Gruppengrößen $n_1 : n_2$ mit drei Stufen (1:1; 1:3 und 3:1) realisiert. Je Verteilung und Stichprobengröße wurden diese Stufen nach folgendem Schema gekreuzt, so daß Studie 1 insgesamt 350 Bedingungen umfaßt.

$$\begin{array}{c}
 \text{Verteilung} \\
 \text{Normal} \\
 \text{Gleich} \\
 g-h-1 \\
 g-h-2 \\
 g-h-3 \\
 g-h-4 \\
 \text{Hyper}
 \end{array}
 \times
 \begin{array}{c}
 N \\
 40 \\
 120
 \end{array}
 \times
 \begin{array}{c}
 \{\sigma_1 : \sigma_2; n_1 : n_2\} \\
 (1:1; 1:1) \\
 (1:1; 1:3) \\
 (3:1; 1:1) \\
 (3:1; 3:1) \\
 (3:1; 1:3)
 \end{array}
 \times
 \begin{array}{c}
 d' \\
 0 \\
 0.4 \\
 0.8 \\
 1.2 \\
 1.6
 \end{array}
 .$$

2.1.5 Ergebnisse der Studie 1

Das Ziel der folgenden Darstellung ist es, möglichst anschaulich und erschöpfend über die Ergebnisse Auskunft zu geben. Daher wurden im wesentlichen graphische Darstellungen gewählt. Im Anhang C finden sich zudem Tabellen auf einem niedrigeren Aggregationsniveau.

2.1.5.1 Punktschätzer

Das Risiko der simulierten Schätzer wird anhand der quadratischen Verlustfunktion beurteilt. Das quadratische Verlustrisiko, auch mittlerer quadratischer Fehler (MSE) genannt, kann in die Varianz und die quadrierte Verzerrung (Bias) des Schätzers aufgespalten werden (Klauer, 1996): $MSE(\hat{\theta}) = V(\hat{\theta}) + B(\hat{\theta})^2$. Je Datensituation wurden für jeden Schätzer folgende (Monte-Carlo-) Schätzungen vorgenommen, wobei auf die Standardfehler dieser simulierten Werte nicht näher eingegangen wird, d.h. es werden keine Konfidenzbänder o.ä. hierfür berechnet:

$$B_{MC}(\hat{\theta}) = E_{MC}(\hat{\theta}) - \theta = \frac{1}{5000} \sum_{m=1}^{5000} \hat{\theta}_m - \theta, \quad MSE_{MC}(\hat{\theta}) = \frac{1}{5000} \sum_{m=1}^{5000} (\hat{\theta}_m - \theta)^2.$$

Da die Vergleichbarkeit der Schätzrisiken aufgrund der Unterschiedlichkeit der Parameterwerte nicht unmittelbar gegeben ist, wurden die von d' verschiedenen Parameter sowie ihre Schätzungen zunächst auf die Skala von d' transformiert, indem mit dem Verhältnis der jeweiligen Streuungen, die zur Standardisierung dienen, multipliziert wurde. Bei der Normalverteilung ist dieser Faktor für d_R identisch mit dem Vorschlag von Algina, Keselman und Penfield (2005a): $\sigma_{w,20\%} / \sigma = 0.642$. Für die anderen Verteilungen wurde analog vorgegangen. Bei Gleichverteilungen lautet der Faktor für d_R also z.B. 0.805 (siehe Tabelle 7, S. 32). Für die Statistiken, bei denen an nur einer Streuung relativiert wird, ergeben sich bei varianzheterogenen Verteilungen entsprechend andere Faktoren, im Falle zweier Normalverteilungen mit Varianzverhältnis neun zu eins lautet der Faktor für $d_R^{S_1}$ z.B. 0.861. Dieses Vorgehen wurde der Betrachtung des relativen Bias $rB(\hat{\theta}) = B(\hat{\theta})/\theta$ und des relativen Fehlers $rMSE(\hat{\theta}) = MSE(\hat{\theta})/\theta^2$ vorgezogen, da es auch für den Fall $\theta = 0$ definiert ist.

Obwohl die so skalierten Simulationsergebnisse der verschiedenen Schätzer prinzipiell vergleichbar sind, bleibt die Betrachtung der über alle Parameter und Verteilungen hinweg gemittelten Verzerrungen und Verlustrisiken problematisch. Denn wie bereits in Kapitel 2.1.4.1 angesprochen, sind die Ergebnisse von der Auswahl der simulierten Verteilungen abhängig. Je größer aber z.B. der Anteil von Verteilungen mit starken Tails ist, desto besser schneiden die robustifizierten Statistiken insgesamt ab. Daher werden im folgenden die Schätzer zunächst getrennt nach ihren Parametern diskutiert und erst im Anschluß an die Auswahl geeigneter Schätzer je Parameter findet eine Gegenüberstellung der herkömmlichen und robustifizierten Statistiken statt.

2.1.5.1.1 Schätzer für d'

Tabelle 10 gibt für die Schätzer für d' die Mittelwerte von B_{MC}^2 und MSE_{MC} insgesamt für alle 350 Simulationsbedingungen sowie getrennt nach Homo- und Heterogenität der Varianzen wieder. Die Unterstreichungen kennzeichnen die besten Werte je Kriterium. Zu beachten ist, daß die Statistiken mit gebündelter Varianz (siehe obere Hälfte von Tabelle 10) im heteroskedastischen Fall, für welchen d nicht definiert ist, auch als Schätzer für d' betrachtet werden. Für dieses Vorgehen spricht, daß die (nichtkorrigierten) Schätzer d_s und d'_s bei gleich großen Gruppen ohnehin dieselben Werte und also auch Erwartungswerte aufweisen.

Tabelle 10. $\overline{B_{MC}^2}$ und $\overline{MSE_{MC}}$ der Schätzer für d' über alle 350 Simulationsbedingungen hinweg sowie getrennt nach Varianzhomogenität und -heterogenität

	Homogenität		Heterogenität		Gesamt	
	$\overline{B_{MC}^2}$	$\overline{MSE_{MC}}$	$\overline{B_{MC}^2}$	$\overline{MSE_{MC}}$	$\overline{B_{MC}^2}$	$\overline{MSE_{MC}}$
d_s	0.004	0.114	0.055	0.187	0.035	0.157
\hat{d}	0.003	0.109	0.051	0.178	0.032	0.150
d_s^{Boot}	0.001	0.107	0.038	0.161	0.023	0.140
\hat{d}^{Boot}	0.001	0.104	0.036	0.155	0.022	0.135
d'_s	0.004	0.111	0.007	0.116	0.006	0.114
\hat{d}'	0.002	0.105	0.005	0.108	0.004	0.106
$d_s'^{Boot}$	0.001	0.104	0.001	0.103	<u>0.001</u>	0.103
\hat{d}'^{Boot}	0.001	0.100	0.002	0.099	0.002	<u>0.099</u>

Anmerkung: Unterstreichungen kennzeichnen die besten Resultate je Kriterium

Wie Tabelle 10 verdeutlicht, weisen die Schätzer mit gebündelter Varianz in heterogenen Situationen einen vergleichsweise großen Bias auf. Doch auch unter den homogenen Bedingungen fällt insgesamt der $\overline{MSE_{MC}}$ nicht kleiner aus als bei den Schätzern mit gemittelter Varianz. Selbst bei Gültigkeit des Modells, die in insgesamt 20 Simulationsbedingungen mit varianzhomogenen Normalverteilungen gegeben ist, schneiden die Schätzer mit gebündelter Varianz nur unwesentlich besser ab als die Schätzer mit gemittelter Varianz; die $\overline{MSE_{MC}}$ lauten hier z.B. 0.087 für \hat{d} und 0.088 für \hat{d}' . Den nur geringen Vorteilen der gebündelten Varianzschätzung bei Gültigkeit des Modells stehen also eventuell erhebliche Nachteile bei Verlassen dieses Modells gegenüber (siehe auch Tabelle C-1 bis Tabelle C-4 im Anhang C). Daher werden die Schätzer mit gebündelter Varianz hier nicht weiter diskutiert. Da in der psychologischen Forschungspraxis die Annahme der Varianzhomogenität nur selten zu begründen ist, wird von der Verwendung dieser Schätzer insgesamt abgeraten. Zu beachten ist hier ferner, daß bei gleich großen Gruppen nur die nichtkorrigierten Schätzer d_s und d'_s dieselben Werte aufweisen. Sofern aber nicht von Varianzhomogenität ausgegangen werden kann, sollte insbesondere bei kleinen Stichproben auch bei gleich großen Gruppen die Korrektur nach Huynh und nicht diejenige nach Hedges gewählt werden (vgl. Huynh, 1989; siehe Tabelle C-1 bis Tabelle C-4 im Anhang C).

Im folgenden werden also nur noch die Schätzer mit gemittelter Varianz betrachtet. Über alle Bedingungen hinweg weisen im Vergleich zu d'_s alle drei korrigierten Schätzer einen etwas geringeren $\overline{B_{MC}^2}$ und einen kleineren $\overline{MSE_{MC}}$ auf (vgl. Tabelle 10), wobei die Effekte der verschiedenen Korrekturen aufgrund der relativ großen Stichprobengrößen eher klein sind. Abbildung 7 gibt nun für alle Bedingungen mit $N = 40$ die Verzerrungen dieser Schätzer wieder. Insgesamt fällt hier auf, daß auch die Bootstrap-Korrekturen über nahezu alle Bedingungen mit $d' > 0$ zu einer Reduktion der Erwartungswerte führen. Ferner wird hier deutlich, daß die Bootstrap-Korrektur des nichtkorrigierten Schätzers $d_s'^{Boot}$ praktisch dieselbe Stärke aufweist wie diejenige der bereits nach Huynh korrigierten Statistik \hat{d}'^{Boot} . Unter allen Verteilungen und Bedingungen zeigt der nichtkorrigierte Schätzer d'_s mit steigender Effektgröße eine zunehmende Überschätzung des Parameters, wobei die deutlichsten Verzerrungen unter den beiden $g-h$ -Verteilungen mit starken Tails ($g-h-1$, $g-h-3$) auftreten. Der größte Bias von d'_s liegt hier unter der $g-h-3$ -Verteilung bei 0.423 für den Fall, daß die kleinere Gruppe die größere Varianz besitzt ($g-h-3$, 1.6; 3:1; 10, 30). Bei den Verteilungen mit normalen oder schwachen Tails (Normal, Gleich, $g-h-4$, Hyper) fallen die Bootstrap-Korrekturen tendentiell zu groß aus und mit steigender Effektgröße kommt es zu einer zunehmenden Unterschätzung des Parameters. Insgesamt weisen die Bootstrap-Korrekturen unter diesen Verteilungen

bei $N = 40$ jeweils einen größeren Bias auf als die Korrektur nach Huynh. Unter den Verteilungen mit starken Tails hingegen verhält es sich umgekehrt, hier zeigen die Bootstrap-Schätzer insgesamt eine deutlich geringere Verzerrung.

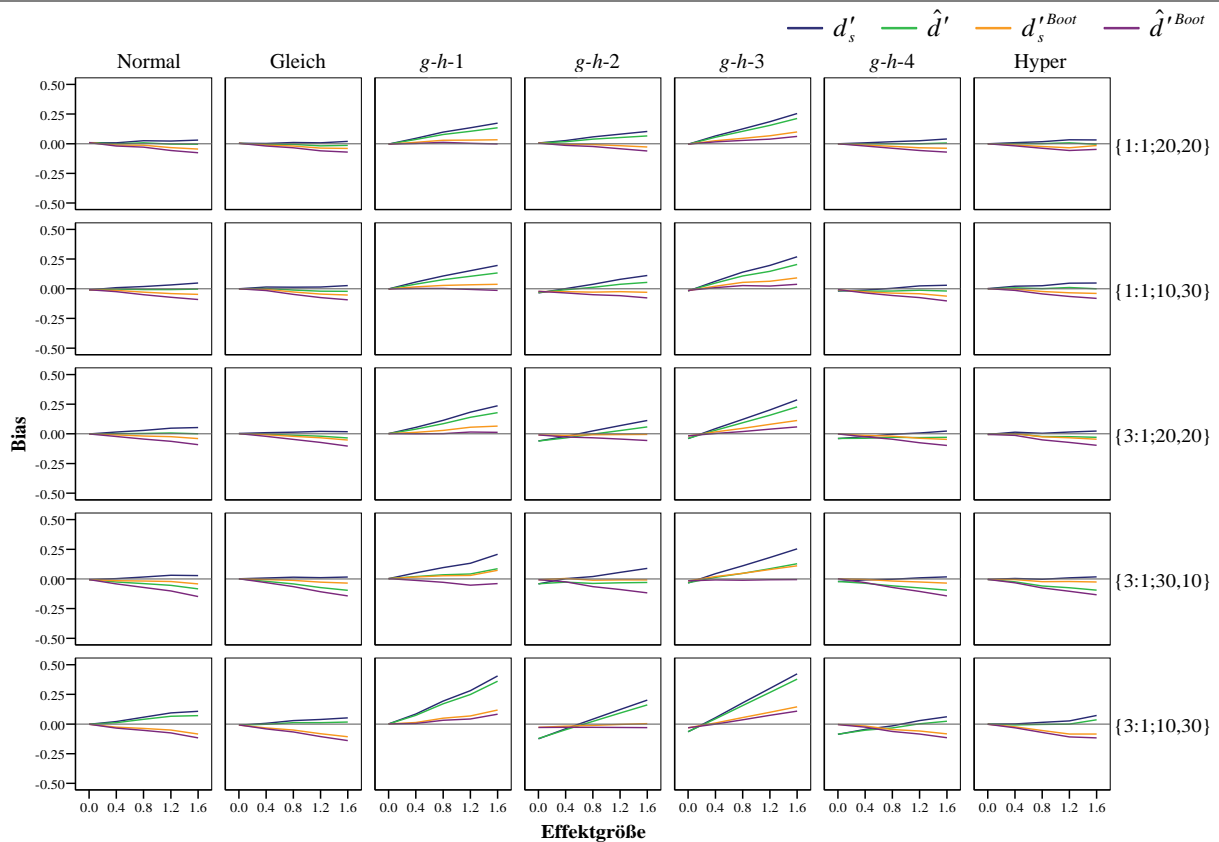


Abbildung 7. Verzerrungen der Schätzer für d'

Über alle Bedingungen und Verteilungen hinweg läßt sich somit kein klar zu bevorzugender Schätzer für d' ausmachen. Um die unter den Verteilungen mit starken Tails auftretenden Überschätzungen zu vermeiden, bieten sich aber ab $N = 40$ die Bootstrap-Schätzer an.

2.1.5.1.2 Schätzer für d^{S_1} und d^{S_2}

Die formal gleichen Schätzer, bei denen an nur einer der beiden Streuungen relativiert wird, werden hier getrennt aufgeführt, da nur rechtsschiefe Verteilungen untersucht wurden. Sofern aber bei schiefen Verteilungen heterogene Varianzen auftreten, unterscheiden sich die Korrelationen zwischen Zähler und Nenner, je nach dem, welche der beiden Streuungen im Nenner gewählt wird. Nur im Falle gleich großer Stichproben aus homogenen, symmetrischen Verteilungen sind die Stichprobenverteilungen der beiden Statistiken gleich. Wie Tabelle 11 deutlich macht, weisen die Bootstrap-Schätzer insgesamt für beide hier betrachteten Parameter sowohl die geringeren Verzerrungen als auch die kleineren \overline{MSE}_{MC} im Vergleich zu den geschlossenen Schätzern auf. Den insgesamt geringsten Bias haben die beiden Bootstrap-Schätzer $d_s^{S_1,Boot}$ und $d_s^{S_2,Boot}$, den geringsten \overline{MSE}_{MC} weisen $\hat{d}^{S_1,Boot}$ und $\hat{d}^{S_2,Boot}$ auf.

Tabelle 11. $\overline{B_{MC}^2}$ und $\overline{MSE_{MC}}$ der Schätzer für d^{S_1} und d^{S_2} über alle 350 Simulationsbedingungen hinweg sowie getrennt nach Varianzhomogenität und -heterogenität

	Homogenität		Heterogenität		Gesamt	
	$\overline{B_{MC}^2}$	$\overline{MSE_{MC}}$	$\overline{B_{MC}^2}$	$\overline{MSE_{MC}}$	$\overline{B_{MC}^2}$	$\overline{MSE_{MC}}$
$d_s^{S_1}$	0.015	0.158	0.011	0.134	0.013	0.144
$\hat{d}_s^{S_1}$	0.007	0.133	0.006	0.117	0.006	0.123
$d_s^{S_1,Boot}$	0.002	0.128	0.002	0.111	<u>0.002</u>	0.118
$\hat{d}_s^{S_1,Boot}$	0.005	0.117	0.004	0.103	0.004	<u>0.109</u>
$d_s^{S_2}$	0.010	0.173	0.018	0.205	0.015	0.192
$\hat{d}_s^{S_2}$	0.006	0.160	0.009	0.180	0.008	0.172
$d_s^{S_2,Boot}$	0.001	0.155	0.002	0.176	<u>0.002</u>	0.168
$\hat{d}_s^{S_2,Boot}$	0.002	0.147	0.004	0.163	0.003	<u>0.157</u>

Anmerkung: Unterstreichungen kennzeichnen die besten Resultate je Kriterium

In Abbildung 8 und Abbildung 9, die wiederum sämtliche Verzerrungen bei $N = 40$ darstellen, wird ein ähnliches Verhalten der verschiedenen Schätzerarten deutlich, wie es sich auch bei der Schätzung von d' gezeigt hat. Über nahezu alle Bedingungen hinweg weisen die nichtkorrigierten Schätzer $d_s^{S_1}$ und $d_s^{S_2}$ größere positive Verzerrungen auf als die Schätzer $\hat{d}_s^{S_1}$ und $\hat{d}_s^{S_2}$; letztere zeigen unter den Verteilungen mit kurzen oder normalen Tails insgesamt die geringsten Verzerrungen. Unter tailstarken Verteilungen hingegen lassen sich bei $\hat{d}_s^{S_1}$ und $\hat{d}_s^{S_2}$ mit steigendem Parameter zunehmend positive Verzerrungen erkennen. Für die meisten Bedingungen mit Parameterwerten größer Null läßt sich eine einheitliche Ordnung der simulierten Erwartungswerte der verschiedenen Schätzer feststellen: Die Bootstrap-Korrektur ist stärker als die t -basierte Korrektur und die Bootstrap-Korrektur für die bereits korrigierte Statistik führt zu einer weiteren Verringerung des Erwartungswertes. Nur unter den schiefen g - h -Verteilungen (g - h -2, g - h -3, g - h -4) treten einige Situationen auf, bei denen diese Ordnung nicht gegeben ist (z.B. g - h -2 mit $\{3:1; 10, 30\}$ in Abbildung 8), was auf die mangelnde Erwartungstreue von $d_s^{S_1}$ und $\hat{d}_s^{S_1}$ für Effektgrößen gleich Null zurückgeht. Diese Verzerrungen bei Effekten von Null können durch das Bootstrap weitgehend korrigiert werden.

Sofern insbesondere eine Überschätzung des Parameters vermieden werden soll, bieten sich die Schätzer $\hat{d}_s^{S_1,Boot}$ und $\hat{d}_s^{S_2,Boot}$ an, die zudem auch die geringsten $\overline{MSE_{MC}}$ aufweisen. Wenn bei diesen Bootstrap-Schätzern jedoch die verwendete Standardabweichung auf nur zehn Werten beruht, führen diese vor allem bei Varianzheterogenität und geringer Tailstärke mit wachsendem Effekt zu einer Unterschätzung des Parameters. Wie die beiden Abbildungen bereits für den Übergang zu einer stabileren Varianzschätzung von $n = 10$ zu $n = 30$ andeuten, so wird anhand der im Anhang C aufgeführten Tabellen deutlich, daß für alle Schätzer die Verzerrungen und Varianzen unter sämtlichen Bedingungen mit zunehmendem N abnehmen (siehe Tabelle C-5 bis Tabelle C-8). Der Vergleich zwischen den Schätzern von d' und d^{S_k} unterstreicht aber das Argument von Hedges und Olkin (1985) gegen die Verwendung der Glass'schen Statistik, denn die Verrechnung beider Gruppenvarianzen führt insgesamt zu präziseren Schätzern mit geringerer Verzerrung (vgl. hierzu Tabelle 10 und Tabelle 11).

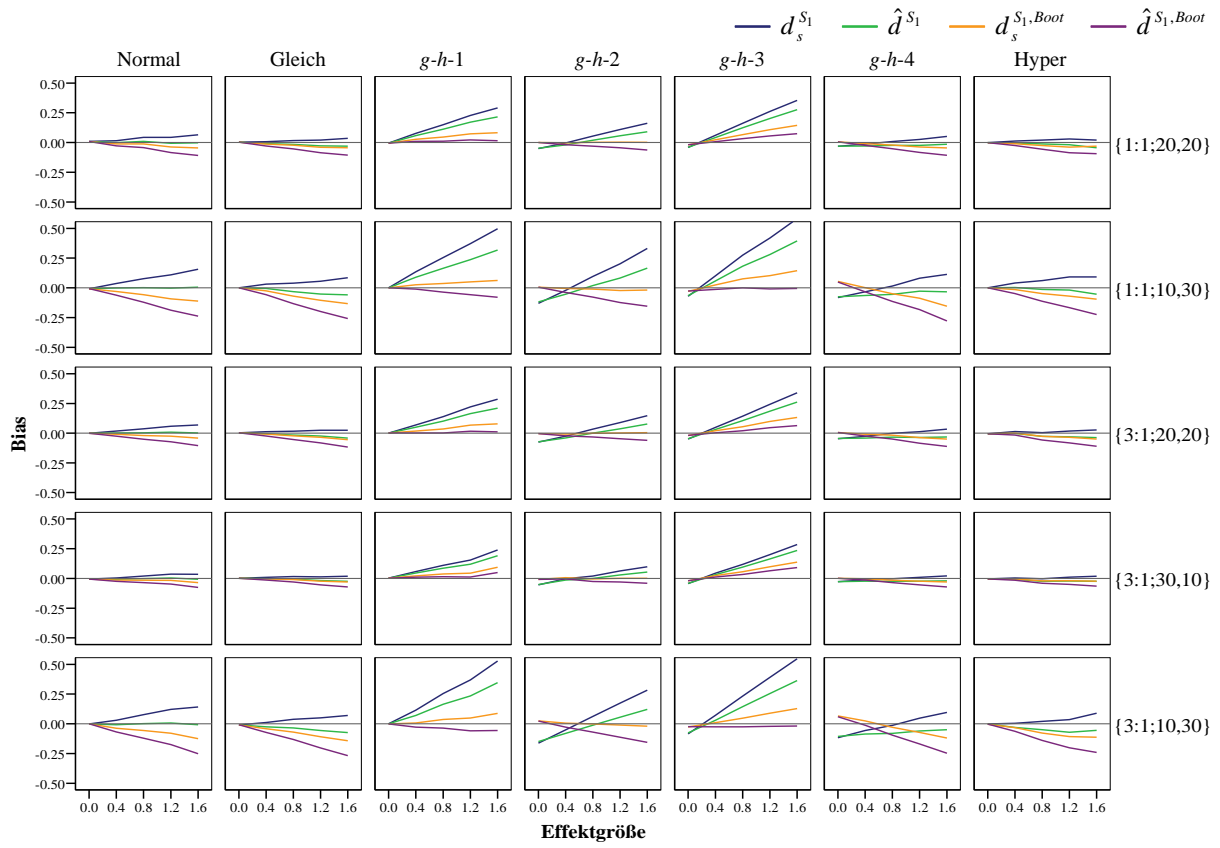


Abbildung 8. Verzerrungen der Schätzer für d^{S_1}

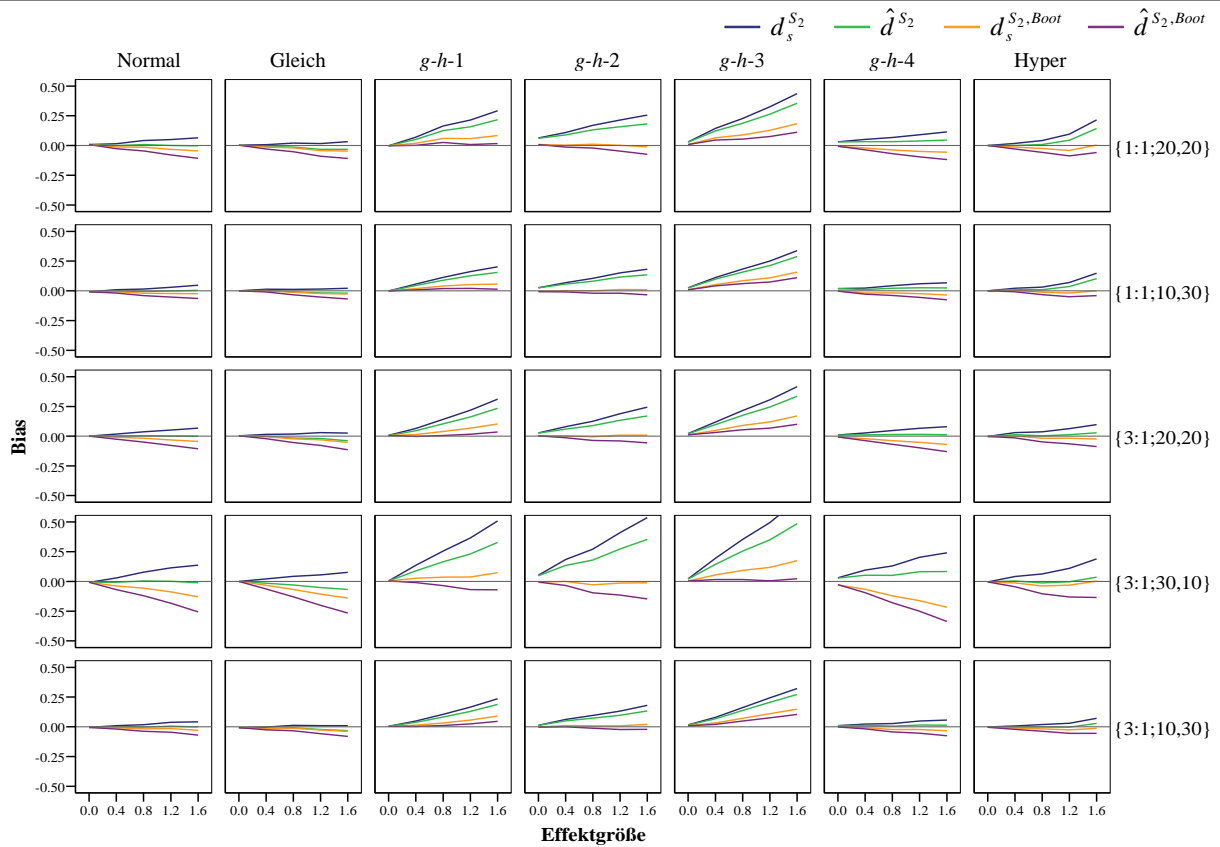


Abbildung 9. Verzerrungen der Schätzer für d^{S_2}

2.1.5.1.3 Schätzer für d'_R

Tabelle 12 führt die gemittelten Verzerrungen und quadratischen Fehler der Schätzer für d'_R insgesamt sowie getrennt nach varianzhomogenen und -heterogenen Bedingungen auf. Da auch hier die Schätzer mit gebündelter Varianz kaum Vorteile bei Varianzhomogenität besitzen (vgl. Tabelle C-9 bis Tabelle C-12), werden diese im folgenden nicht näher betrachtet. Von den Schätzern mit gemittelter Varianz weisen die Bootstrap-Korrekturen über alle Bedingungen hinweg die kleinsten Verzerrungen auf. Anders als bei der Schätzung von d' ist der quadratische Fehler dieser Verfahren jedoch höher als der Fehler der t -basierten Korrektur nach Huynh.

Tabelle 12. $\overline{B_{MC}^2}$ und $\overline{MSE_{MC}}$ der Schätzer für d'_R über alle 350 Simulationsbedingungen hinweg sowie getrennt nach Varianzhomogenität und -heterogenität

	Homogenität		Heterogenität		Gesamt	
	$\overline{B_{MC}^2}$	$\overline{MSE_{MC}}$	$\overline{B_{MC}^2}$	$\overline{MSE_{MC}}$	$\overline{B_{MC}^2}$	$\overline{MSE_{MC}}$
$d'_{s,R}$	0.005	0.138	0.061	0.271	0.038	0.218
\hat{d}'_R	0.002	0.128	0.053	0.251	0.033	0.202
$d'^{Boot}_{s,R}$	0.003	0.156	0.046	0.284	0.029	0.233
\hat{d}'^{Boot}_R	0.002	0.146	0.041	0.265	0.025	0.217
$d'_{s,R}$	0.003	0.134	0.008	0.192	0.006	0.169
\hat{d}'_R	0.001	0.120	0.005	0.171	0.003	<u>0.150</u>
$d'^{tBoot}_{s,R}$	0.002	0.152	0.002	0.213	<u>0.002</u>	0.188
\hat{d}'^{tBoot}_R	0.000	0.138	0.003	0.194	<u>0.002</u>	0.172

Anmerkung: Unterstreichungen kennzeichnen die besten Resultate je Kriterium

Die Effekte der verschiedenen Korrekturen sowie das Verhalten unter den einzelnen Bedingungen sei auch hier für $N = 40$ anhand von Abbildung 10 verdeutlicht. Anders als bei den KQ-Schätzern unterscheiden sich die Eigenschaften der robustifizierten Schätzer kaum zwischen den verschiedenen zugrundeliegenden Verteilungen. Die Verzerrungen fallen unter den Verteilungen mit starken Tails nicht größer aus. Während das Verhalten der KQ-Schätzer unter den negativen hypergeometrischen Modellen vergleichbar mit demjenigen unter der Normalverteilung ist (Abbildung 7), nehmen die diskreten Verteilungen hier eine Sonderstellung ein, denn die Kurvenverläufe unterscheiden sich grundsätzlich von denen unter den stetigen Verteilungen. Da die negativen hypergeometrischen Verteilungen nicht verschoben werden, ändern sich mit der Variation der Effektgröße auch die Eigenschaften dieser Verteilungen. Zu vermuten ist, daß hier vor allem die Anzahl der Bindungen (Ties) zusammen mit dem Trimmanteil von 20% zu diesen nicht monotonen Verläufen führen (Abbildung 10).

Unter allen Bedingungen zeigt der nichtkorrigierte Schätzer $d'_{s,R}$ den höchsten und in aller Regel positiv verzerrten Erwartungswert, wobei die stärksten Überschätzungen bei Gegenläufigkeit von Gruppengröße und -varianz auftreten. Zumeist führt die Korrektur nach Huynh zu einer weniger verzerrten Schätzung. Ausnahmen lassen sich jedoch bei unbalancierten, heterogenen Situationen mit größerer Varianz in der größeren Gruppe beobachten, bei welchen die Korrektur zu einer deutlichen Unterschätzung des Parameters führt. Eine ähnliche Beziehung läßt sich zwischen den beiden Bootstrap-Korrekturen beobachten, denn auch hier liegt der Erwartungswert des Verfahrens mit der korrigierten Statistik stets unterhalb des Verfahrens mit unkorrigierter Statistik. Unter Varianzhomogenität oder bei gleich großen Gruppen weist das Verfahren \hat{d}'^{Boot}_R die geringsten Verzerrungen auf. Nur unter dem negativen hypergeometrischen Modell und unter der Gleichverteilung

lung treten bei größeren Parameterwerten deutlichere Unterschätzungen auf. Wie im Falle von d' kann auch für die Schätzung von d'_R die Bootstrap-Korrektur in Verbindung mit dem bereits korrigierten Schätzer $\hat{d}'_{R,Boot}$ empfohlen werden, wenn eine Überschätzung des Parameters vermieden werden soll. Als problematisch erweist sich dieses Verfahren jedoch bei ungleich großen Gruppen, wenn die größere Gruppe zudem eine größere Varianz aufweist. Dann nimmt die Unterschätzung des Parameters mit steigendem Effekt unter allen Verteilungen zu. Als Alternative zu diesem Verfahren bietet sich die Bootstrap-Korrektur $d'^{Boot}_{s,R}$ an, die bei insgesamt gleicher quadratischer Verzerrung zu einer etwas stärkeren Überschätzung größerer Parameterwerte führt.

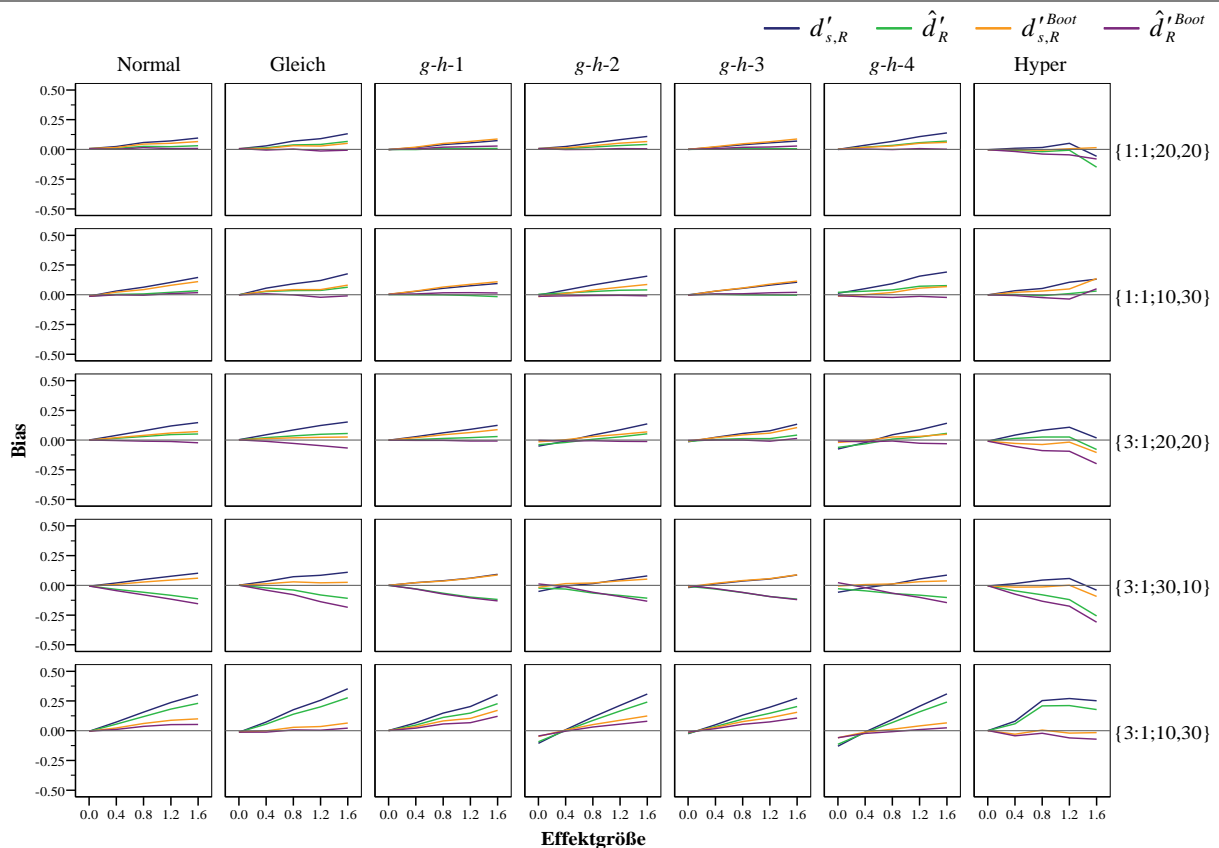


Abbildung 10. Verzerrungen der Schätzer für d'_R

2.1.5.1.4 Schätzer für $d^{S_1}_R$ und $d^{S_2}_R$

Tabelle 13 faßt die Eigenschaften der Schätzer für $d^{S_1}_R$ und $d^{S_2}_R$ über alle Bedingungen hinweg zusammen. Hier wird bereits anhand der extremen MSE_{MC} -Werte deutlich, daß die Anwendbarkeit von Bootstrap-Verfahren an eine Mindestgröße der Stichprobe gebunden ist. Während bei der Schätzung von d^{S_k} ein $n = 10$ in der Gruppe, die zur Schätzung der Standardabweichung dient, für die Konstruktion einer geeigneten Bootstrap-Verteilung noch hinreichend groß ist, so ist dies bei einem Trimmanteil von 20% auf beiden Seiten nicht mehr der Fall. Abbildung 11 und Abbildung 12 geben wiederum die Verzerrungen für alle Bedingungen mit $N = 40$ wieder. Sofern also die Varianzschätzung anhand der kleineren Gruppe mit $n = 10$ vorgenommen wird, kommt es zu einer drastischen Unterschätzung des Parameters. Dieses Versagen der Bootstrap-Biaskorrektur ist durch das Auftreten von Bootstrap-Stichproben mit Varianzen nahe Null zu erklären, die zu extremen Schätzwerten in der Bootstrap-Verteilung führen.

Tabelle 13. $\overline{B_{MC}^2}$ und $\overline{MSE_{MC}}$ der Schätzer für $d_R^{S_1}$ und $d_R^{S_2}$ über alle 350 Simulationsbedingungen hinweg sowie getrennt nach Varianzhomogenität und -heterogenität

	Homogenität		Heterogenität		Gesamt	
	$\overline{B_{MC}^2}$	$\overline{MSE_{MC}}$	$\overline{B_{MC}^2}$	$\overline{MSE_{MC}}$	$\overline{B_{MC}^2}$	$\overline{MSE_{MC}}$
$d_{s,R}^{S_1}$	0.025	0.284	0.015	0.297	0.019	0.291
$\hat{d}_R^{S_1}$	0.003	0.203	0.002	0.221	<u>0.002</u>	<u>0.214</u>
$d_{s,R}^{S_1,Boot}$	0.084	11.512	0.053	8.083	0.066	9.455
$\hat{d}_R^{S_1,Boot}$	0.101	9.686	0.066	7.502	0.080	8.376
$d_{s,R}^{S_2}$	0.010	0.223	0.025	0.296	0.019	0.267
$\hat{d}_R^{S_2}$	0.003	0.196	0.006	0.229	<u>0.005</u>	<u>0.216</u>
$d_{s,R}^{S_2,Boot}$	0.002	0.254	0.086	7.547	0.053	4.630
$\hat{d}_R^{S_2,Boot}$	0.002	0.230	0.090	7.076	0.055	4.338

Anmerkung: Unterstreichungen kennzeichnen die besten Resultate je Kriterium

Während unter symmetrischen Verteilungen und bei Parameterwerten von Null diese Extremwerte auf beiden Seiten gleichermaßen auftreten, kommt es mit einem Anstieg des Parameters häufiger zu extremen Bootstrap-Statistiken auf einer Seite der Bootstrap-Verteilung, wodurch die Bias-Korrektur verzerrt wird. Im hier realisierten Algorithmus werden nur Bootstrap-Stichproben verworfen, deren Effektgröße nicht definiert ist, d.h. deren Standardabweichung gleich Null ist. Sofern zumindest 20 Werte für die Schätzung der Standardabweichung genutzt werden, schneidet die Bootstrap-Korrektur analog zu den Ergebnissen für die Schätzung von d'_R ab.

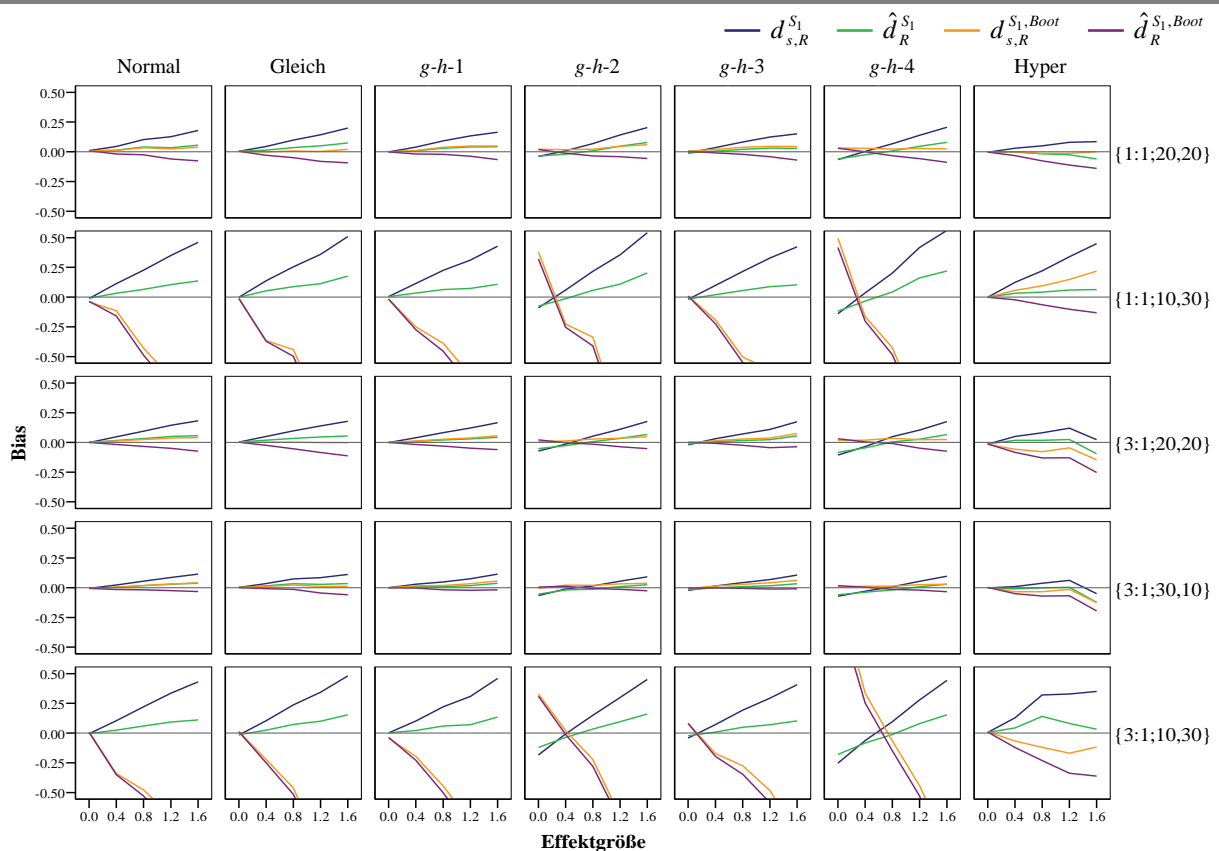


Abbildung 11. Verzerrungen der Schätzer für $d_R^{S_1}$

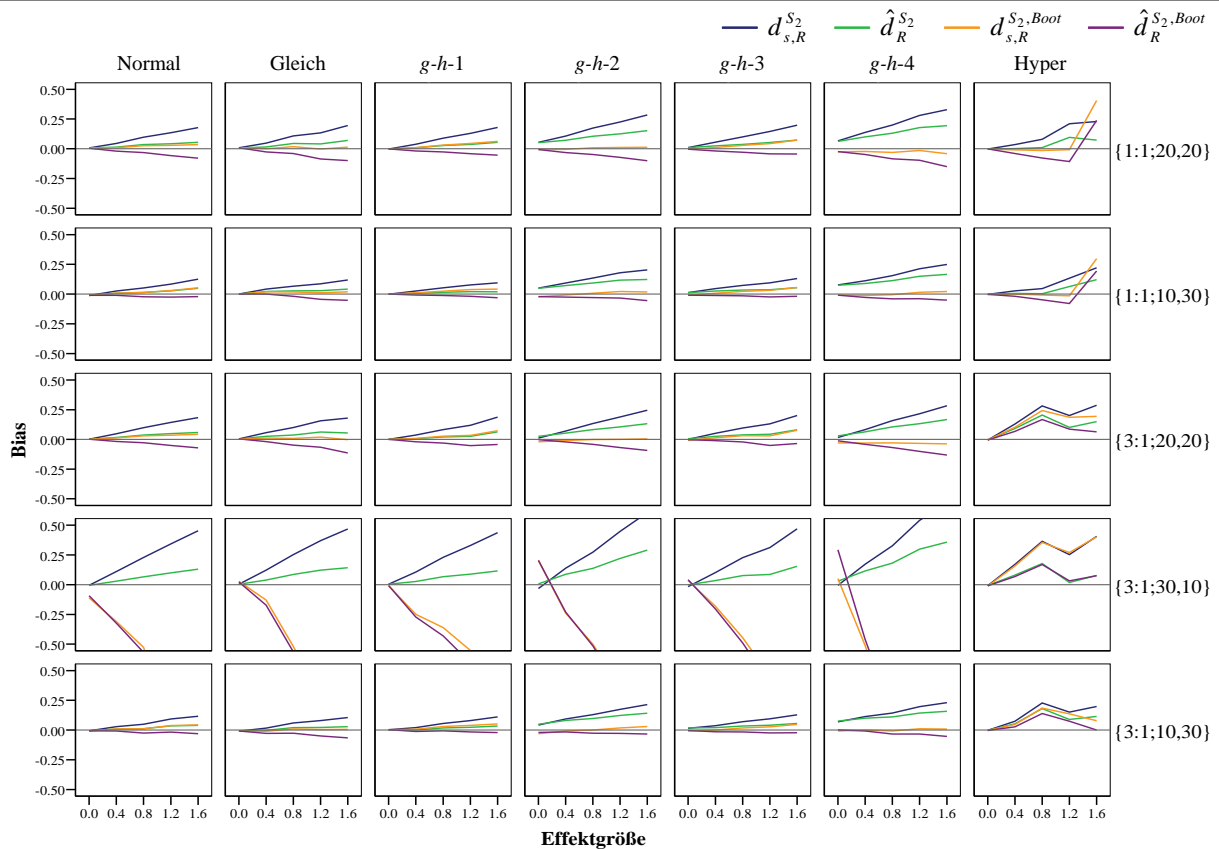


Abbildung 12. Verzerrungen der Schätzer für $d_R^{S_2}$

2.1.5.1.5 Vergleich von KQ- und robustifizierter Schätzung

Bis auf die beschriebenen Probleme der Bootstrap-Biaskorrektur bei zu kleinen Stichproben, weisen Tabelle 10 bis Tabelle 13 sowie Abbildung 7 bis Abbildung 12 darauf hin, daß die Verzerrungen der korrigierten herkömmlichen und der robustifizierten Schätzer ähnlich groß sind. Wie in diesen Tabellen an den über alle Bedingungen hinweg gemittelten MSE_{MC} abzulesen ist, gelingt die Schätzung der herkömmlichen Statistiken insgesamt mit höherer Präzision. Da die Verzerrungen jedoch von den Verteilungsmodellen abhängen, sollte der Vergleich zwischen den Schätzerarten auch getrennt nach den Verteilungsmodellen vorgenommen werden.

Tabelle 14. $\overline{B^2}_{MC}$ und \overline{MSE}_{MC} von \hat{d}' und \hat{d}'_R getrennt nach Verteilungen

	\hat{d}'		\hat{d}'_R	
	$\overline{B^2}_{MC}$	\overline{MSE}_{MC}	$\overline{B^2}_{MC}$	\overline{MSE}_{MC}
Normal	0.001	0.099	0.003	0.146
Gleich	0.000	0.092	0.004	0.207
<i>g-h-1</i>	0.009	0.152	0.003	0.096
<i>g-h-2</i>	0.002	0.085	0.003	0.090
<i>g-h-3</i>	0.014	0.144	0.002	0.078
<i>g-h-4</i>	0.001	0.073	0.003	0.128
Hyper	0.000	0.100	0.006	0.308
Gesamt	0.004	0.106	0.003	0.150

Am Beispiel der beiden Schätzer \hat{d}' und \hat{d}'_R sind in Tabelle 14 die gemittelten Verzerrungs- und Fehlerquadrate aufgeführt. Hier wird nun deutlich, daß \hat{d}' unter Verteilungen mit kurzen oder normalen Tails sowohl einen kleineren Bias als auch eine geringere Varianz aufweist, während dies bei Verteilungen mit starken Tails umgekehrt ist. Da in der vorliegenden Arbeit mehr Modelle mit kurzen bzw. normalen Tails simuliert wurden, fällt die Präzision insgesamt für \hat{d}' günstiger aus als für \hat{d}'_R .

2.1.5.2 Intervallschätzer

Als wesentliche Kriterien zur Beurteilung der Robustheit werden die folgenden drei Kennwerte herangezogen:

- $pCP_{.936}^{.966}$ gibt die relative Häufigkeit an, mit welcher die simulierten Überdeckungswahrscheinlichkeiten $1 - \hat{\alpha}^*$ bei der gewählten nominalen Wahrscheinlichkeit von $1 - \alpha = .95$ zwischen .936 und .966 liegen (vgl. Kap. 2.1.4.4);
- $pCP_{.936}^{1.000}$ ist die entsprechende Häufigkeit für den Bereich .936 bis 1;
- $|CP - .95|$ notiert die mittlere absolute Abweichung des simulierten vom nominalen Konfidenzniveau.

Die Verteilung der simulierten Überdeckungswahrscheinlichkeiten wird zudem durch Angabe des arithmetischen Mittels (Mittel), des Minimums (Min) und des Maximums (Max) gekennzeichnet. Als nachgeordnetes Kriterium dient schließlich die Länge der Intervalle (hier gleichbedeutend mit Intervallbreite), die mit dem arithmetischen Mittel (Mittel) und dem Maximum (Max) beschrieben wird. Die Vergleichbarkeit der Intervalllängen wurde über die einheitliche Skalierung der Intervallgrenzen erreicht, wie sie auch für die Punktschätzungen eingesetzt wurde und in Kapitel 2.1.5.1 beschrieben wird.

2.1.5.2.1 Schätzer für d'

Tabelle 15 faßt die Ergebnisse zur Überdeckung der Intervallschätzer für d' über alle 350 Simulationsbedingungen hinweg dar. In der oberen Hälfte sind die Ergebnisse zu den sieben Schätzern auf Basis der Statistiken mit gebündelter Varianz, in der unteren Hälfte die sieben Schätzer mit nicht gebündelter Varianz eingetragen. Vergleicht man zunächst jeweils die gleich konstruierten Schätzer miteinander, so wird deutlich, daß die Schätzverfahren auf Basis der Statistiken mit nicht gebündelter Varianz hinsichtlich der Überdeckungskriterien $pCP_{.936}^{.966}$, $pCP_{.936}^{1.000}$ und $|CP - .95|$ besser abschneiden als die Verfahren mit gebündelter Varianz. Die jeweils günstigsten Werte sind unterstrichen. Am Beispiel der beiden nichtzentralen Intervalle $NC(d_s)$ und $NC(d'_s)$ werden die Unterschiede zwischen diesen beiden Verfahrensgruppen in Abbildung 13 für alle Bedingungen mit $N = 40$ verdeutlicht.

Die wesentlichen Vorteile des Welch-Verfahrens $NC(d'_s)$ zeigen sich insbesondere, wenn die kleinere Gruppe die größere Varianz besitzt (siehe untere Reihe in Abbildung 13). In diesen Situationen unterschreitet $NC(d_s)$ sehr deutlich die nominale Überdeckungswahrscheinlichkeit über alle Verteilungen und Parameterwerte hinweg. Sofern die größere Gruppe auch die größere Varianz hat, liegt die simulierte Überdeckungswahrscheinlichkeit von $NC(d_s)$ hingegen tendentiell über der nominalen. Nur unter den tailstarken Verteilungen $g-h-1$ und $g-h-3$ erweist sich dies bei steigender Effektgröße als Vorteil gegenüber der Methode nach Welch, da die Überdeckungswahrscheinlichkeiten beider Verfahren abnehmen.

Tabelle 15. Simulierte Überdeckungswahrscheinlichkeiten und Intervallängen der Schätzer für d' über alle 350 Simulationsbedingungen

	Überdeckung						Länge	
	$pCP_{.936}^{.966}$	$pCP_{.936}^{1.000}$	$ CP - .95 $	Mittel	Min	Max	Mittel	Max
NC (d_s)	.391	.591	.073	.894	.322	.997	1.138	1.847
Perc (d_s)	.277	.280	.068	.883	.334	.970	1.238	3.121
BCA (d_s)	.540	.551	.045	.909	.409	.972	1.217	2.978
Perc-t (d_s)	.369	.454	.055	.901	.375	.986	1.179	2.609
Perc (\hat{d})	.280	.283	.066	.884	.350	.970	1.220	3.059
BCA (\hat{d})	.551	.563	.045	.908	.372	.972	1.199	2.918
Perc-t (\hat{d})	.389	.460	.057	.899	.344	.987	1.162	2.557
NC (d'_s)	.589	<u>.780</u>	.025	.941	.729	.999	1.171	2.418
Perc (d'_s)	.369	.371	.042	.908	.671	.970	1.200	2.855
BCA (d'_s)	<u>.697</u>	.711	.019	.936	.815	.974	1.172	2.571
Perc-t (d'_s)	.474	.609	.028	.932	.814	.994	1.143	2.372
Perc (\hat{d}')	.389	.391	.037	.913	.688	.970	1.169	2.780
BCA (\hat{d}')	.689	<u>.697</u>	<u>.018</u>	.936	.820	.974	1.142	2.504
Perc-t (\hat{d}')	.449	.569	.029	.931	.816	.994	1.115	2.314

Anmerkung: Unterstreichungen kennzeichnen die besten Resultate je Kriterium

Sofern ein Effekt von Null vorliegt, entsprechen diese Ergebnisse den seit langem bekannten Befunden zur α -Robustheit des t -Tests und seinem antikonservativen bzw. konservativen Verhalten in diesen Datensituationen (Büning, 1991, S.130 ff.).

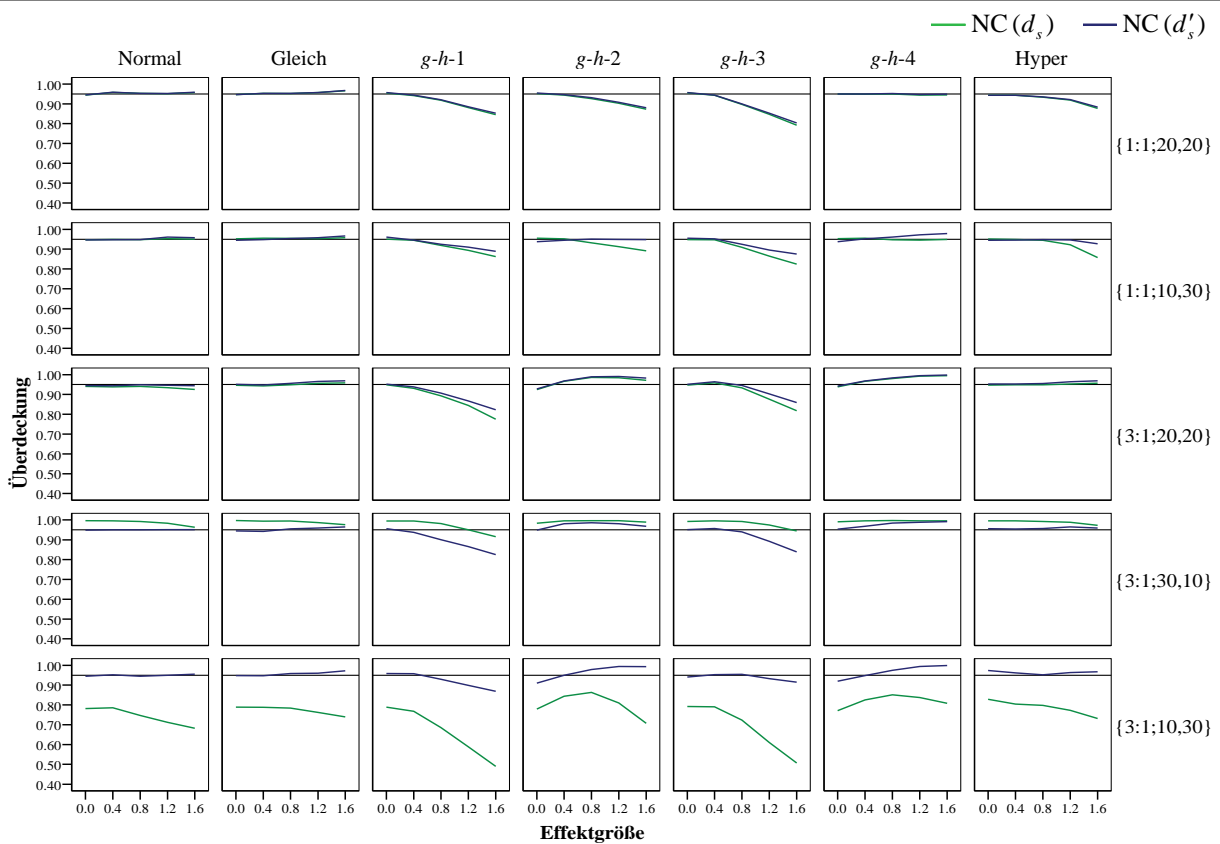


Abbildung 13. Simulierte Überdeckungswahrscheinlichkeiten von $NC(d_s)$ und $NC(d'_s)$

Bei gleich großen Gruppen sind die Unterschiede zwischen $NC(d'_s)$ und $NC(d'_s)$ vergleichsweise gering. Vorteile der Freiheitsgradenkorrektur ergeben sich hier bei Varianzheterogenität nur unter den tailstarken Verteilungen $g-h-1$ und $g-h-3$, wobei diese relativ gering sind. Bei gleich großen, varianzhomogenen Gruppen zeigen sich keine nennenswerten Unterschiede zwischen den beiden Verfahren (vgl. Tabelle C-17 bis Tabelle C-20 im Anhang C). Da sich dieses Verhalten, nämlich unwesentliche Vorteile bei Varianzhomogenität, aber z.T. erhebliche Nachteile bei Heterogenität, auch bei den Bootstrap-Verfahren findet (vgl. Anhang C), werden die Verfahren, bei denen die Varianzen gebündelt werden bzw. keine Freiheitsgradenkorrektur stattfindet, im folgenden nicht näher betrachtet.

Wie durch Tabelle 15 angedeutet wird, sind die Effekte der Verwendung des Korrekturfaktors nach Huynh auf die Überdeckung der Bootstrap-Verfahren insgesamt jeweils gering. Um einen graphischen Vergleich der Bootstrap-Verfahren mit dem nichtzentralen Intervall zu ermöglichen, werden in Abbildung 14 nur die drei Bootstrap-Verfahren ohne Korrektur dem nichtzentralen Intervall gegenübergestellt.

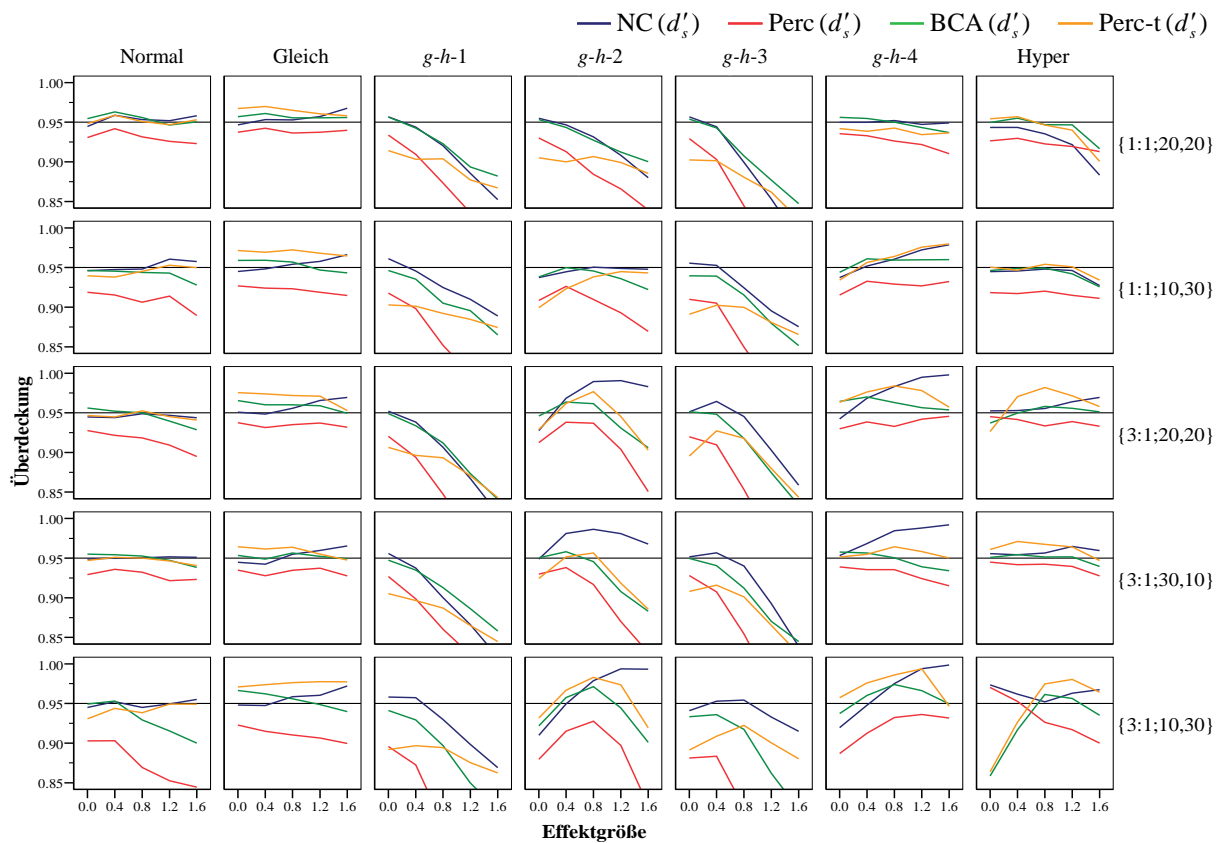


Abbildung 14. Simulierte Überdeckungswahrscheinlichkeiten ausgewählter Schätzer für d'

Ein erster Blick auf die Abbildung 14 verdeutlicht, daß die Überdeckung von d' zum einen bei Verteilungen mit stärkeren Tails ($g-h-1$, $g-h-2$ und $g-h-3$), und zum anderen bei Gegenläufigkeit von Gruppengröße und Varianz (vgl. untere Reihe in Abbildung 14) weniger gut gelingt. Hinsichtlich einzelner Verfahren fällt vor allem auf, daß das Perzentil-Bootstrap in nahezu allen Bedingungen die nominale Überdeckungswahrscheinlichkeit unterschreitet und insgesamt deutlich schlechter abschneidet als die anderen Verfahren. Daher fällt sowohl das abgebildete Perzentil-Verfahren mit

dem nichtkorrigierten Punktschätzer als auch das Perzentil-Bootstrap mit korrigiertem Punktschätzer aus der weiteren Betrachtung heraus, für weitere Details siehe Anhang C. Hinsichtlich der anderen drei Verfahren sticht insbesondere noch das Perzentil- t -Bootstrap durch seine mangelnde Überdeckung bei Effektgrößen von Null unter den beiden tailstarken Verteilungen $g-h-1$ und $g-h-3$ heraus (siehe hierzu auch die vergleichsweise schlechten Eigenschaften, wie sie in Tabelle 15 oben deutlich werden). Bei Gegenläufigkeit von Gruppengröße und Varianz ist dieses Verhalten auch für das BCA-Verfahren unter der negativen hypergeometrischen Verteilung zu beobachten. Insgesamt aber weist die graphische Analyse bei $N = 40$ das nichtzentrale sowie das BCA-Verfahren als die beiden besten Methoden zur Intervallschätzung von d' aus.

Um zusätzlich die Bedingungen mit $N = 120$ in die Betrachtung mit einzubeziehen, werden die Eigenschaften beider Methoden – die Bootstrap-Variante mit korrigiertem Schätzer wird nun ebenfalls berücksichtigt – zusätzlich getrennt nach Verteilung und Stichprobengröße in tabellarischer Form aufgeführt (vgl. Tabelle 16). Da zwischen den Parameterwerten der Effektgröße und dem Verhältnis der Varianzen nicht getrennt wird, liegen den hier angegebenen Statistiken jeweils 25 Simulationsbedingungen zugrunde. Unter der Normalverteilung schneidet das aus diesem Modell entwickelte NC-Verfahren im Vergleich mit den Bootstrap-Verfahren am besten ab, es kann hier unter allen Bedingungen als robust bezeichnet werden. Abgesehen von der Normalverteilung schneidet das NC-Verfahren ferner bei $N = 40$ unter der $g-h-1$ - und der $g-h-3$ -Verteilung hinsichtlich der Überdeckungswahrscheinlichkeit etwas besser ab als die Bootstrap-Verfahren. Unter allen anderen Bedingungen jedoch ist das Verhalten der Bootstrap-Verfahren insgesamt als günstiger zu beurteilen. Wesentlicher Unterschied zwischen dem NC-Verfahren und dem Bootstrap ist, daß sich nur für die Bootstrap-Intervalle die Robustheit mit steigendem Stichprobenumfang verbessert. Für die beiden Bootstrap-Verfahren sind in nahezu allen Fällen – einzige Ausnahme ist die Verteilung $g-h-3$, bei welcher die Asymptotik deutlich langsamer zu sein scheint – die Anteile der robusten Intervalle ($pCP_{.936}^{.966}$) bei $N = 120$ höher als bei $N = 40$, und die absolute Abweichung von der nominalen Überdeckung ($|\overline{CP} - .95|$) ist dann in aller Regel geringer. Da sich bei $N = 40$ keine deutlichen Vorteile des NC-Verfahrens erkennen lassen, kann mit den genannten Einschränkungen hier eine Empfehlung zur Verwendung des BCA-Verfahrens getroffen werden. Während sich die Überdeckungswahrscheinlichkeiten von $BCA(d'_s)$ und $BCA(\hat{d}')$ kaum unterscheiden, kann eine Entscheidung zwischen diesen beiden Verfahren recht eindeutig aufgrund der durchgängig geringeren Intervallängen für $BCA(\hat{d}')$ getroffen werden.

Will man also eine einfache Empfehlung aus den durchgeführten Simulationen zur Schätzung von Konfidenzintervallen für d' ableiten, so würde die Wahl auf $BCA(\hat{d}')$ fallen. Zu beachten ist jedoch, daß die Überdeckung eines solchen Intervalls aufgrund der Verwendung von herkömmlichen Schätzern mit zunehmender Tailstärke der zugrundeliegenden Population schlechter ausfällt, also die nominale Wahrscheinlichkeit zunehmend unterschreitet.

Tabelle 16. Simulierte Überdeckungswahrscheinlichkeiten und Intervallängen ausgewählter Schätzer für d' getrennt nach Verteilungen

			Überdeckung					Länge		
			$pCP_{.936}^{.966}$	$pCP_{.936}^{1.000}$	$ CP - .95 $	Mittel	Min	Max	Mittel	Max
Normal	40	NC (d'_s)	1.000	1.000	.004	.950	.944	.961	1.487	2.273
		BCA (d'_s)	.800	.800	.010	.944	.900	.963	1.524	2.228
		BCA (\hat{d}'_s)	.720	.720	.012	.941	.902	.963	1.473	2.179
	120	NC (d'_s)	1.000	1.000	.005	.951	.940	.965	0.847	1.269
		BCA (d'_s)	.920	.920	.005	.948	.932	.959	0.848	1.219
		BCA (\hat{d}'_s)	.960	.960	.005	.948	.932	.960	0.839	1.211
Gleich	40	NC (d'_s)	.840	1.000	.008	.956	.942	.972	1.484	2.244
		BCA (d'_s)	.960	1.000	.007	.955	.940	.966	1.516	2.140
		BCA (\hat{d}'_s)	.960	.960	.008	.951	.917	.966	1.465	2.094
	120	NC (d'_s)	.720	1.000	.009	.958	.946	.975	0.847	1.266
		BCA (d'_s)	1.000	1.000	.004	.953	.945	.961	0.824	1.125
		BCA (\hat{d}'_s)	1.000	1.000	.004	.951	.943	.961	0.815	1.117
$g-h-1$	40	NC (d'_s)	.400	.400	.042	.911	.823	.961	1.510	2.418
		BCA (d'_s)	.240	.240	.045	.905	.815	.957	1.612	2.571
		BCA (\hat{d}'_s)	.280	.280	.041	.909	.820	.957	1.556	2.504
	120	NC (d'_s)	.280	.280	.063	.889	.774	.958	0.853	1.309
		BCA (d'_s)	.320	.320	.035	.916	.846	.954	0.951	1.541
		BCA (\hat{d}'_s)	.320	.320	.033	.918	.849	.954	0.942	1.532
$g-h-2$	40	NC (d'_s)	.360	.800	.024	.956	.880	.994	1.485	2.278
		BCA (d'_s)	.560	.600	.020	.935	.883	.971	1.396	2.076
		BCA (\hat{d}'_s)	.520	.560	.021	.935	.885	.970	1.348	2.022
	120	NC (d'_s)	.400	.840	.024	.959	.863	.995	0.847	1.274
		BCA (d'_s)	.600	.600	.014	.939	.900	.965	0.766	1.056
		BCA (\hat{d}'_s)	.640	.640	.014	.939	.903	.964	0.758	1.048
$g-h-3$	40	NC (d'_s)	.520	.520	.035	.918	.803	.964	1.514	2.414
		BCA (d'_s)	.360	.360	.048	.903	.817	.954	1.567	2.372
		BCA (\hat{d}'_s)	.360	.360	.043	.907	.822	.952	1.513	2.319
	120	NC (d'_s)	.400	.400	.061	.892	.729	.965	0.856	1.320
		BCA (d'_s)	.360	.360	.044	.906	.826	.954	0.917	1.414
		BCA (\hat{d}'_s)	.360	.360	.042	.909	.829	.954	0.908	1.405
$g-h-4$	40	NC (d'_s)	.440	.960	.021	.966	.920	.998	1.478	2.221
		BCA (d'_s)	.840	.960	.010	.954	.934	.974	1.359	2.143
		BCA (\hat{d}'_s)	.840	.920	.011	.951	.906	.974	1.312	2.089
	120	NC (d'_s)	.400	1.000	.023	.972	.945	.999	0.846	1.262
		BCA (d'_s)	1.000	1.000	.005	.954	.938	.964	0.733	1.056
		BCA (\hat{d}'_s)	.960	.960	.006	.952	.931	.964	0.725	1.049
Hyper	40	NC (d'_s)	.720	.840	.013	.950	.883	.973	1.488	2.259
		BCA (d'_s)	.800	.800	.012	.942	.859	.961	1.537	2.157
		BCA (\hat{d}'_s)	.720	.720	.013	.939	.859	.960	1.486	2.110
	120	NC (d'_s)	.760	.880	.012	.949	.862	.976	0.847	1.264
		BCA (d'_s)	1.000	1.000	.004	.951	.939	.961	0.858	1.125
		BCA (\hat{d}'_s)	1.000	1.000	.004	.950	.938	.962	0.850	1.118

2.1.5.2.2 Schätzer für d^{S_1} und d^{S_2}

Tabelle 17 führt die über alle 350 Simulationsbedingungen hinweg aggregierten Charakteristika der Überdeckungswahrscheinlichkeiten aller simulierten Schätzer für d^{S_1} und d^{S_2} auf. Eine Zusammenfassung der Schätzungen beider Parameter je Versuchsbedingung ist zunächst nicht sinnvoll, da nicht alle Kombinationen von Varianzverhältnis, Balance, Effektgröße und Verteilungsschiefe realisiert wurden. Zudem macht nur eine getrennte Betrachtung die Auswirkungen der Schiefe von Verteilungen auf das Überdeckungsverhalten dieser Schätzer deutlich. Daß also die Schätzer für d^{S_1} insgesamt bessere Überdeckungen zeigen als die Schätzer für d^{S_2} ist allein dem Design der Studie geschuldet. Die Unterstreichungen kennzeichnen jeweils getrennt für die geschätzten Parameter den besten Wert der jeweiligen Statistik. Für die mittlere und die maximale Überdeckung erscheint eine solche Beurteilung nicht angebracht. Insgesamt deutet sich für beide Parameter eine relativ gute Überdeckung der BCA-Bootstraps an.

Tabelle 17. Simulierte Überdeckungswahrscheinlichkeiten und Intervallängen der Schätzer für d^{S_1} und d^{S_2} über alle 350 Simulationsbedingungen

	Überdeckung						Länge	
	$pCP_{.936}^{.966}$	$pCP_{.936}^{1.000}$	$ CP - .95 $	Mittel	Min	Max	Mittel	Max
NC ($d_s^{S_1}$)	.503	<u>.817</u>	.028	.948	.757	.999	1.246	2.650
Perc ($d_s^{S_1}$)	.380	.383	.048	.902	.633	.970	1.363	3.992
BCA ($d_s^{S_1}$)	<u>.677</u>	.726	.022	.935	.811	.988	1.279	3.133
Perc-t ($d_s^{S_1}$)	.463	.651	.028	.937	.804	.998	1.199	2.858
Perc ($\hat{d}_s^{S_1}$)	.426	.429	.040	.910	.704	.970	1.298	3.649
BCA ($\hat{d}_s^{S_1}$)	.669	.700	<u>.021</u>	.935	.818	.985	1.218	2.818
Perc-t ($\hat{d}_s^{S_1}$)	.411	.580	.030	.932	.809	.996	1.144	2.613
NC ($d_s^{S_2}$)	.403	.449	.047	.907	.647	.983	1.241	2.520
Perc ($d_s^{S_2}$)	.317	.320	.043	.907	.619	.970	1.516	4.899
BCA ($d_s^{S_2}$)	<u>.546</u>	<u>.569</u>	.026	.928	.772	.977	1.430	3.446
Perc-t ($d_s^{S_2}$)	.449	.469	.037	.917	.768	.980	1.395	3.103
Perc ($\hat{d}_s^{S_2}$)	.349	.351	.037	.913	.686	.970	1.462	4.479
BCA ($\hat{d}_s^{S_2}$)	.531	.549	<u>.025</u>	.928	.804	.975	1.382	3.164
Perc-t ($\hat{d}_s^{S_2}$)	.403	.423	.039	.914	.777	.980	1.350	3.022

Anmerkung: Unterstreichungen kennzeichnen die besten Resultate je Kriterium

Da sich auch hier die analog konstruierten Bootstrap-Verfahren, ob nun mit oder ohne Korrektur des Punktschätzers, ähnlich verhalten, wird der Übersichtlichkeit halber jeweils nur eines dieser Bootstrap-Verfahren für die Bedingungen mit $N = 40$ graphisch dargestellt. Abbildung 15 zeigt die Überdeckungswahrscheinlichkeiten der Schätzer für d^{S_1} , Abbildung 16 gibt die Schätzer für d^{S_2} wieder.

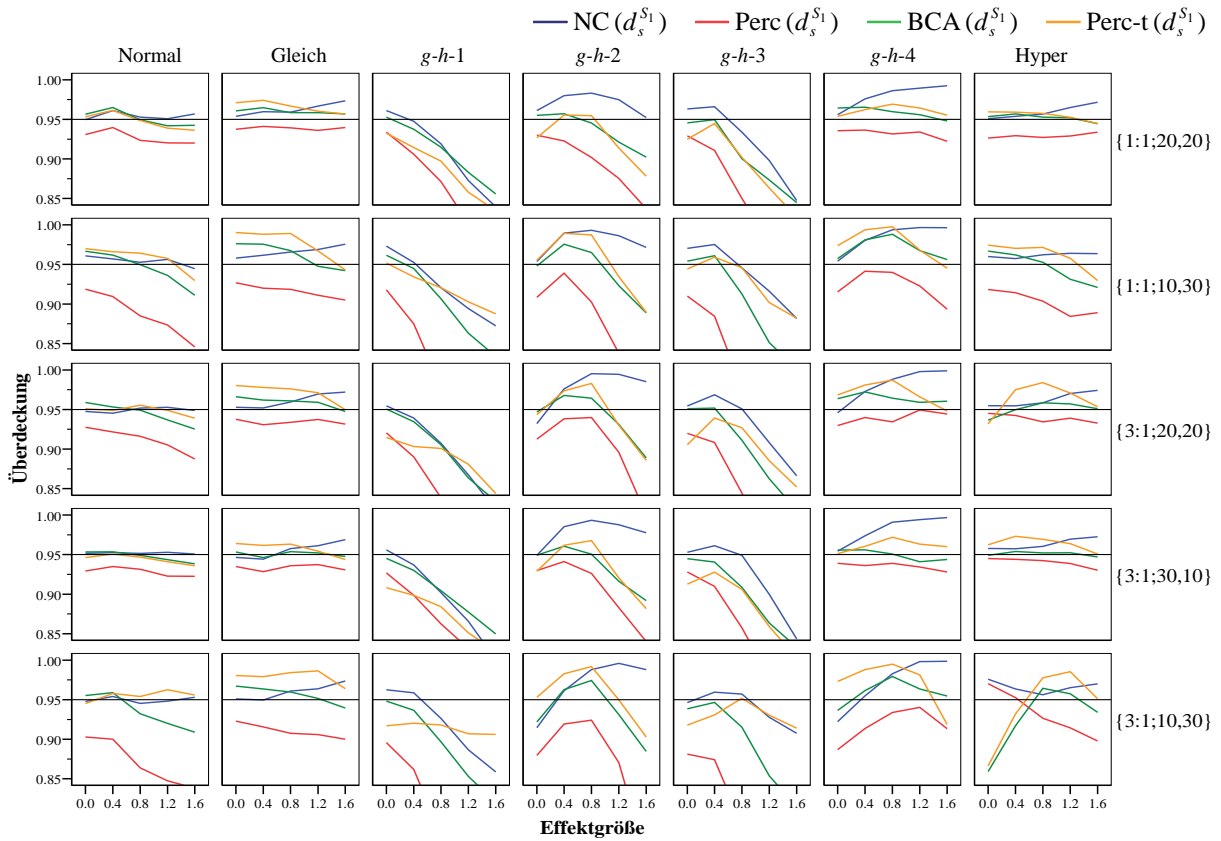


Abbildung 15. Simulierte Überdeckungswahrscheinlichkeiten ausgewählter Schätzer für d^{S_1}

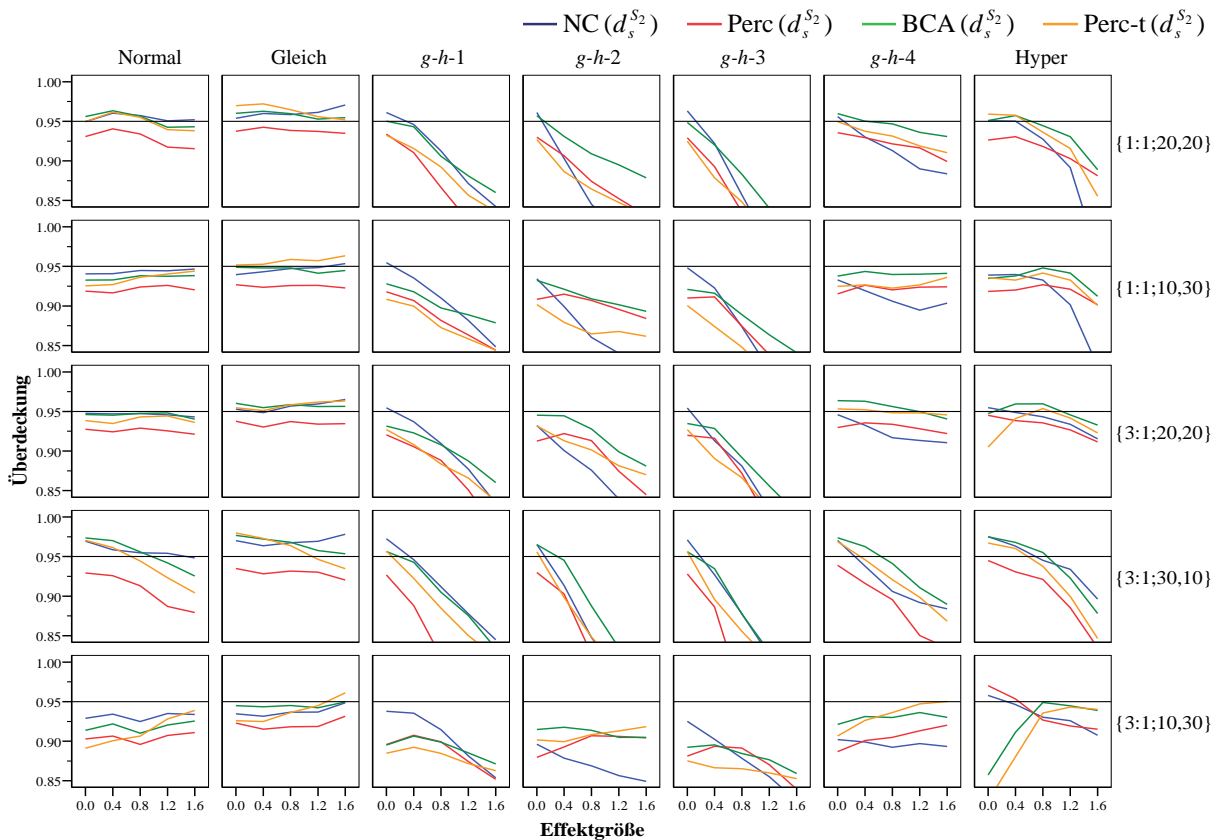


Abbildung 16. Simulierte Überdeckungswahrscheinlichkeiten ausgewählter Schätzer für d^{S_2}

Zunächst sei darauf hingewiesen, daß die Schätzer für d^{S_1} und d^{S_2} nur bei den Bedingungen mit homogenen Varianzen und gleich großen Gruppen direkt vergleichbar sind, da sich alle anderen Simulationsbedingungen entweder durch die Größe der (Nenner-)Varianz oder die Größe der Gruppe, an deren Varianz standardisiert wird, unterscheiden. Betrachtet man nun die jeweils oberste Reihe der beiden Abbildungen im Vergleich, so fällt auf, daß die Verfahren für d^{S_1} und d^{S_2} nur bei symmetrischen Verteilungen (Normal, Gleich, $g-h-1$) auch zu den gleichen Ergebnissen führen. Unter den schiefen Verteilungen ist dies hingegen außer bei einer Effektgröße von Null nicht der Fall. Als Ursache kann hierfür die Korreliertheit von Mittelwert und Varianz bei schiefen Verteilungen angeführt werden, weshalb auch die erforderliche Unabhängigkeit von Zähler und Nenner in der t -Statistik bzw. Effektgröße nicht mehr gegeben ist. Bei den hier betrachteten rechtsschiefen Verteilungen ergeben sich für die Schätzer für d^{S_1} positive Korrelationen zwischen der Differenz der Mittelwerte und der Varianz, bei d^{S_2} hingegen sind diese negativ. Die Ergebnisse deuten darauf hin, daß die Verfahren bei positiver Korrelation von Zähler und Nenner tendentiell besser überdecken als bei negativer. Insbesondere unter $g-h-2$ und $g-h-3$ führt eine Zunahme des Effekts bei der Schätzung von d^{S_2} schneller zu einer Abnahme der Überdeckungswahrscheinlichkeit als bei der Schätzung von d^{S_1} .

Über alle Bedingungen hinweg fällt im Vergleich der Bootstrap-Verfahren wiederum das Perzentil-Bootstrap durch seine durchgängig deutlichste Unterschreitung der nominalen Überdeckungswahrscheinlichkeit auf, weshalb dieses Verfahren aus der weiteren Betrachtung herausgenommen wird (Details siehe Tabelle C-21 bis Tabelle C-24 im Anhang). Die Unterschiede zwischen dem Perzentil- t - und dem BCA-Verfahren sind für die Schätzungen von d^{S_2} recht deutlich, unter den meisten Bedingungen liegt die beobachtete Überdeckungswahrscheinlichkeit des BCA-Verfahrens näher an der nominalen. Für d^{S_1} gibt es hingegen einige Situationen, in denen das BCA-Verfahren schlechter abschneidet. Hierzu gehören insbesondere die Bedingungen unter den tailstarken Verteilungen $g-h-1$ und $g-h-3$, wenn die zur Standardisierung herangezogene Varianz auf nur $n_1 = 10$ Werten beruht. Da das BCA-Verfahren aber in den meisten Fällen eine bessere oder ähnliche Überdeckung zeigt – für d^{S_1} liegt die Überdeckungswahrscheinlichkeit des BCA-Verfahrens insgesamt in 73% aller Bedingungen näher an der nominalen, für d^{S_2} sind es 75% –, und dies auch durch die Gesamtanalyse (vgl. Tabelle 17) für beide Parameter gestützt wird, fällt auch hier das Perzentil- t -Verfahren aus der näheren Analyse heraus. Obwohl das NC-Verfahren insgesamt keine bessere Überdeckung aufweist als das Perzentil- t -Verfahren, wird dieses im folgenden dennoch in die detaillierte Betrachtung mit einbezogen, da es eine methodische Alternative zum Bootstrap darstellt, die auch ohne Kenntnis der Primärdaten anhand von Kennwerten einer Stichprobe zur Schätzung von Intervallen herangezogen werden kann.

Die nach Verteilungen und Stichprobengröße getrennte Analyse ergibt zunächst, daß sich das Verhalten der drei ausgewählten Verfahren zur Schätzung von d^{S_1} und d^{S_2} am ehesten unter der Gleichverteilung als robust bezeichnen läßt. Die mittleren absoluten Abweichungen liegen hier bei höchstens .012, der Anteil robuster Intervalle ($pCP_{.936}^{1.000}$) liegt bei mindestens .880 und die minimale Überdeckung bei .921. Wie Tabelle 18 und Tabelle 19 zudem deutlich machen, bietet das NC-Verfahren nur selten Vorteile gegenüber den beiden BCA-Verfahren. Hinsichtlich der mittleren absoluten Abweichung von der nominalen Überdeckung weist das NC-Verfahren nur bei $N = 40$ für die Schätzung von d^{S_1} unter der Normal-, der Gleich- und der negativen hypergeometrischen Verteilung geringfügig bessere Werte auf als die BCA-Verfahren. Unter allen anderen Bedingungen und insbesondere bei $N = 120$ schneiden die beiden BCA-Verfahren zum Teil erheblich besser ab (vgl. z.B. $N = 120$ $g-h-1$, $g-h-2$, $g-h-3$ in Tabelle 19).

Tabelle 18. Simulierte Überdeckungswahrscheinlichkeiten und Intervallängen ausgewählter Schätzer für d^{S_1} getrennt nach Verteilungen

			Überdeckung						Länge	
			$pCP_{.936}^{.966}$	$pCP_{.936}^{1.000}$	$ CP - .95 $	Mittel	Min	Max	Mittel	Max
Normal	40	NC ($d_s^{S_1}$)	1.000	1.000	.004	.952	.945	.961	1.571	2.361
		BCA ($d_s^{S_1}$)	.760	.800	.012	.945	.909	.967	1.686	2.478
		BCA ($\hat{d}_s^{S_1}$)	.680	.680	.014	.942	.901	.965	1.582	2.251
	120	NC ($d_s^{S_1}$)	1.000	1.000	.003	.951	.941	.958	0.878	1.296
		BCA ($d_s^{S_1}$)	.920	.920	.006	.949	.932	.963	0.892	1.282
		BCA ($\hat{d}_s^{S_1}$)	.920	.920	.007	.948	.928	.963	0.876	1.249
Gleich	40	NC ($d_s^{S_1}$)	.680	1.000	.012	.961	.944	.976	1.550	2.307
		BCA ($d_s^{S_1}$)	.800	1.000	.010	.958	.939	.976	1.643	2.311
		BCA ($\hat{d}_s^{S_1}$)	.800	.880	.011	.952	.921	.975	1.544	2.115
	120	NC ($d_s^{S_1}$)	.640	1.000	.011	.961	.947	.978	0.874	1.288
		BCA ($d_s^{S_1}$)	1.000	1.000	.004	.953	.947	.961	0.850	1.160
		BCA ($\hat{d}_s^{S_1}$)	1.000	1.000	.004	.951	.938	.961	0.834	1.130
$g-h-1$	40	NC ($d_s^{S_1}$)	.360	.400	.047	.909	.817	.973	1.677	2.649
		BCA ($d_s^{S_1}$)	.320	.320	.051	.900	.820	.962	1.928	3.133
		BCA ($\hat{d}_s^{S_1}$)	.320	.320	.047	.904	.827	.960	1.806	2.818
	120	NC ($d_s^{S_1}$)	.200	.200	.078	.874	.770	.961	0.908	1.368
		BCA ($d_s^{S_1}$)	.360	.360	.044	.908	.839	.959	1.058	1.710
		BCA ($\hat{d}_s^{S_1}$)	.360	.360	.041	.911	.845	.959	1.038	1.664
$g-h-2$	40	NC ($d_s^{S_1}$)	.200	.920	.029	.975	.915	.996	1.620	2.446
		BCA ($d_s^{S_1}$)	.440	.560	.024	.937	.885	.976	1.573	2.492
		BCA ($\hat{d}_s^{S_1}$)	.440	.560	.024	.937	.884	.977	1.473	2.281
	120	NC ($d_s^{S_1}$)	.280	.960	.026	.973	.930	.996	0.888	1.316
		BCA ($d_s^{S_1}$)	.520	.600	.020	.935	.874	.968	0.771	1.143
		BCA ($\hat{d}_s^{S_1}$)	.560	.600	.018	.936	.882	.968	0.756	1.109
$g-h-3$	40	NC ($d_s^{S_1}$)	.440	.560	.030	.930	.843	.975	1.696	2.650
		BCA ($d_s^{S_1}$)	.400	.400	.053	.899	.814	.961	1.854	2.959
		BCA ($\hat{d}_s^{S_1}$)	.400	.400	.048	.903	.818	.960	1.735	2.668
	120	NC ($d_s^{S_1}$)	.360	.440	.060	.896	.757	.968	0.917	1.389
		BCA ($d_s^{S_1}$)	.400	.400	.052	.898	.811	.953	0.993	1.596
		BCA ($\hat{d}_s^{S_1}$)	.400	.400	.048	.903	.829	.953	0.975	1.552
$g-h-4$	40	NC ($d_s^{S_1}$)	.200	.960	.032	.980	.922	.999	1.560	2.313
		BCA ($d_s^{S_1}$)	.800	1.000	.013	.960	.936	.988	1.459	2.527
		BCA ($\hat{d}_s^{S_1}$)	.760	.920	.012	.955	.934	.985	1.369	2.312
	120	NC ($d_s^{S_1}$)	.200	1.000	.032	.982	.946	.999	0.875	1.287
		BCA ($d_s^{S_1}$)	1.000	1.000	.006	.955	.941	.966	0.709	1.085
		BCA ($\hat{d}_s^{S_1}$)	.960	1.000	.006	.953	.939	.966	0.696	1.057
Hyper	40	NC ($d_s^{S_1}$)	.720	1.000	.013	.963	.951	.976	1.553	2.321
		BCA ($d_s^{S_1}$)	.760	.800	.012	.945	.859	.967	1.635	2.321
		BCA ($\hat{d}_s^{S_1}$)	.800	.800	.014	.940	.858	.965	1.536	2.125
	120	NC ($d_s^{S_1}$)	.760	1.000	.011	.961	.949	.978	0.875	1.285
		BCA ($d_s^{S_1}$)	1.000	1.000	.004	.952	.939	.962	0.857	1.146
		BCA ($\hat{d}_s^{S_1}$)	.960	.960	.004	.950	.934	.961	0.841	1.117

Tabelle 19. Simulierte Überdeckungswahrscheinlichkeiten und Intervallängen ausgewählter Schätzer für d^{S_2} getrennt nach Verteilungen

			Überdeckung						Länge	
			$pCP_{.936}^{.966}$	$pCP_{.936}^{1.000}$	$ CP - .95 $	Mittel	Min	Max	Mittel	Max
Normal	40	NC ($d_s^{S_2}$)	.760	.800	.008	.946	.925	.969	1.544	1.985
		BCA ($d_s^{S_2}$)	.600	.680	.015	.941	.910	.973	1.608	2.055
		BCA ($\hat{d}_s^{S_2}$)	.520	.560	.017	.938	.908	.973	1.537	1.868
	120	NC ($d_s^{S_2}$)	1.000	1.000	.003	.950	.943	.957	0.866	1.087
		BCA ($d_s^{S_2}$)	.960	.960	.005	.948	.933	.961	0.875	1.091
		BCA ($\hat{d}_s^{S_2}$)	.960	.960	.006	.947	.925	.961	0.863	1.082
Gleich	40	NC ($d_s^{S_2}$)	.720	.920	.011	.954	.932	.978	1.521	1.917
		BCA ($d_s^{S_2}$)	.880	1.000	.008	.954	.941	.977	1.513	1.790
		BCA ($\hat{d}_s^{S_2}$)	.880	.960	.007	.951	.926	.975	1.448	1.731
	120	NC ($d_s^{S_2}$)	.760	1.000	.009	.957	.943	.983	0.862	1.088
		BCA ($d_s^{S_2}$)	1.000	1.000	.003	.953	.946	.959	0.828	1.026
		BCA ($\hat{d}_s^{S_2}$)	1.000	1.000	.003	.951	.945	.959	0.817	1.018
g-h-1	40	NC ($d_s^{S_2}$)	.280	.320	.047	.906	.834	.973	1.667	2.357
		BCA ($d_s^{S_2}$)	.160	.160	.049	.901	.834	.956	1.961	2.902
		BCA ($\hat{d}_s^{S_2}$)	.160	.160	.045	.906	.847	.955	1.872	2.625
	120	NC ($d_s^{S_2}$)	.240	.240	.073	.878	.756	.957	0.901	1.170
		BCA ($d_s^{S_2}$)	.160	.160	.041	.909	.844	.953	1.069	1.560
		BCA ($\hat{d}_s^{S_2}$)	.160	.160	.038	.912	.852	.953	1.054	1.518
g-h-2	40	NC ($d_s^{S_2}$)	.080	.080	.086	.866	.767	.965	1.646	2.379
		BCA ($d_s^{S_2}$)	.200	.200	.043	.909	.806	.965	2.160	3.446
		BCA ($\hat{d}_s^{S_2}$)	.200	.200	.042	.910	.819	.964	2.063	3.146
	120	NC ($d_s^{S_2}$)	.160	.160	.088	.862	.738	.953	0.886	1.138
		BCA ($d_s^{S_2}$)	.360	.360	.024	.927	.872	.962	1.137	1.705
		BCA ($\hat{d}_s^{S_2}$)	.360	.360	.023	.928	.876	.962	1.121	1.660
g-h-3	40	NC ($d_s^{S_2}$)	.120	.160	.082	.871	.753	.971	1.714	2.520
		BCA ($d_s^{S_2}$)	.080	.080	.069	.881	.772	.956	2.121	3.276
		BCA ($\hat{d}_s^{S_2}$)	.120	.120	.062	.888	.804	.956	2.024	2.967
	120	NC ($d_s^{S_2}$)	.200	.200	.125	.826	.647	.962	0.917	1.225
		BCA ($d_s^{S_2}$)	.080	.080	.061	.890	.791	.960	1.181	1.784
		BCA ($\hat{d}_s^{S_2}$)	.080	.080	.058	.893	.809	.960	1.164	1.736
g-h-4	40	NC ($d_s^{S_2}$)	.120	.160	.039	.913	.884	.970	1.556	2.076
		BCA ($d_s^{S_2}$)	.680	.720	.015	.941	.890	.974	1.885	2.953
		BCA ($\hat{d}_s^{S_2}$)	.520	.560	.017	.939	.880	.973	1.801	2.708
	120	NC ($d_s^{S_2}$)	.240	.240	.026	.925	.893	.958	0.867	1.091
		BCA ($d_s^{S_2}$)	1.000	1.000	.005	.951	.939	.963	0.970	1.288
		BCA ($\hat{d}_s^{S_2}$)	1.000	1.000	.005	.950	.937	.963	0.956	1.255
Hyper	40	NC ($d_s^{S_2}$)	.440	.480	.030	.925	.781	.975	1.563	2.031
		BCA ($d_s^{S_2}$)	.560	.640	.020	.936	.857	.975	1.741	3.294
		BCA ($\hat{d}_s^{S_2}$)	.560	.640	.021	.934	.857	.974	1.667	3.164
	120	NC ($d_s^{S_2}$)	.520	.520	.028	.923	.770	.958	0.870	1.099
		BCA ($d_s^{S_2}$)	.920	.920	.007	.949	.915	.965	0.968	1.637
		BCA ($\hat{d}_s^{S_2}$)	.920	.920	.007	.949	.911	.965	0.955	1.617

Der detaillierte Vergleich bestätigt also das globale Ergebnis in Tabelle 17, daß als generelle Empfehlung eines der BCA-Verfahren in Frage kommt, wobei auch hier wiederum das Verfahren mit korrigiertem Punktschätzer bei ähnlicher Überdeckungsleistung zu durchgängig kürzeren Intervallen

len führt. Auch hier ist diese Empfehlung jedoch mit dem Hinweis zu versehen, daß die tatsächliche Überdeckungswahrscheinlichkeit unter Verteilungen mit starken Tails deutlich unterhalb der nominalen Wahrscheinlichkeit liegt.

2.1.5.2.3 Schätzer für d'_R

Wie bei den KQ-Schätzern zeigt sich auch bei den robustifizierten Statistiken, daß die Verfahren mit gebündelten Varianzen insgesamt deutlich schlechtere Überdeckungswahrscheinlichkeiten aufweisen (vgl. Tabelle 20). Da die Unterschiede zwischen gebündelten und gemittelten Varianzschätzern bei der Intervallschätzung von d' bereits dargestellt wurden und die Verfahren mit der Welch-Statistik auch bei gleich großen Gruppen einen geringeren absoluten Fehler haben als die Verfahren mit gebündelter Varianz, werden diese Verfahren hier nicht näher betrachtet und für Detailergebnisse sei wiederum auf den Anhang C (Tabelle C-25 und Tabelle C-26) verwiesen.

Tabelle 20. Simulierte Überdeckungswahrscheinlichkeiten und Intervalllängen der Schätzer für d'_R über alle 350 Simulationsbedingungen

	Überdeckung						Länge	
	$pCP_{.936}^{.966}$	$pCP_{.936}^{1.000}$	$ \overline{CP} - .95 $	Mittel	Min	Max	Mittel	Max
NC($d_{s,R}$)	.346	.531	.068	.898	.430	.999	1.234	2.636
Perc($d_{s,R}$)	.663	.771	.025	.934	.631	.982	1.565	6.445
BCA($d_{s,R}$)	.671	.671	.028	.923	.616	.961	1.555	6.187
Perc-t($d_{s,R}$)	.380	.514	.042	.918	.506	.999	1.436	4.947
Perc(\hat{d}_R)	.640	.740	.025	.934	.645	.986	1.524	6.223
BCA(\hat{d}_R)	.674	.674	.028	.922	.596	.961	1.514	5.973
Perc-t(\hat{d}_R)	.371	.494	.045	.914	.480	.999	1.398	4.777
NC($d'_{s,R}$)	.671	.789	.014	.946	.795	.993	1.298	3.341
Perc($d'_{s,R}$)	.783	<u>.986</u>	<u>.009</u>	.956	.931	.987	1.559	6.035
BCA($d'_{s,R}$)	<u>.837</u>	.837	<u>.009</u>	.942	.833	.959	1.556	5.758
Perc-t($d'_{s,R}$)	.540	.640	.017	.941	.819	.998	1.412	4.333
Perc(\hat{d}'_R)	.803	.980	<u>.009</u>	.956	.909	.988	1.478	5.687
BCA(\hat{d}'_R)	.786	.786	.010	.941	.833	.959	1.475	5.436
Perc-t(\hat{d}'_R)	.460	.551	.022	.935	.787	.997	1.347	4.139

Anmerkung: Unterstreichungen kennzeichnen die besten Resultate je Kriterium

Anders als bei den Verfahren mit KQ-Statistiken deuten sich hier über alle Bedingungen hinweg hinsichtlich der Überdeckung die beiden Perzentil-Bootstraps als die robustesten Verfahren an. Insbesondere zeigt Perc($d'_{s,R}$) mit simulierten Überdeckungswahrscheinlichkeiten zwischen .931 und .987 ein sehr robustes Verhalten. Bezüglich des mittleren absoluten Fehlers ($|\overline{CP} - .95|$) bieten die beiden BCA-Verfahren insgesamt eine vergleichbare Robustheit. Hinsichtlich der durchschnittlichen Intervalllängen fällt auf, daß die BCA-Verfahren gegenüber den Perzentil-Verfahren keine Vorteile bieten. Die Intervalle des NC-Verfahrens fallen durchschnittlich kürzer aus, überdecken jedoch nicht so gut wie die genannten Bootstrap-Verfahren. Auch hier sei zunächst das Überdeckungsverhalten für $N = 40$ anhand von Abbildung 17 untersucht.

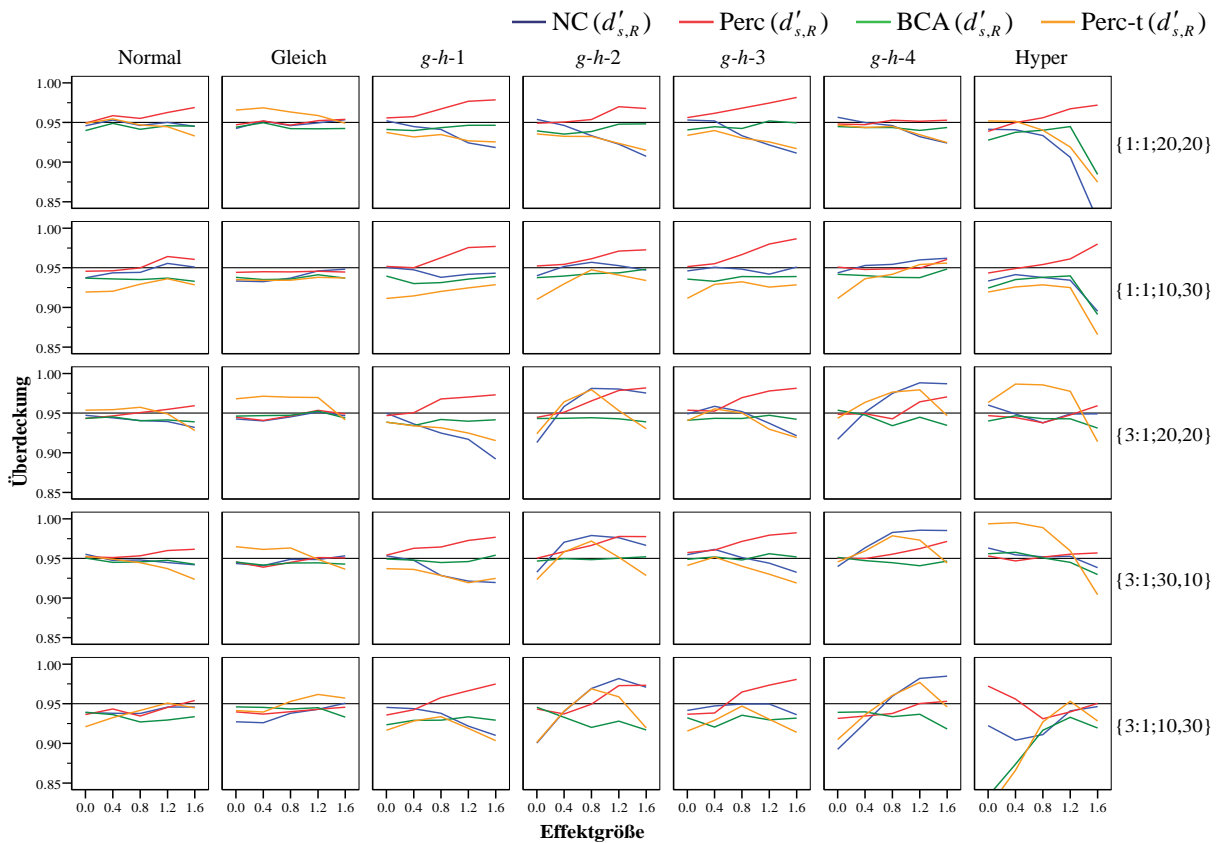


Abbildung 17. Simulierte Überdeckungswahrscheinlichkeiten ausgewählter Schätzer für d'_R

Im Vergleich zu den herkömmlichen Schätzern (siehe Abbildung 14) wird deutlich, daß die Überdeckung insgesamt mit den robustifizierten Statistiken unter den Verteilungen mit schwachen Tails ähnlich gut und unter Verteilungen mit starken Tails besser gelingt. Im Einzelnen fällt auf, daß das Perzentil-Verfahren mit robustifizierten Effektgrößen in aller Regel bei steigender Effektgröße mit einer Zunahme der Überdeckungswahrscheinlichkeit reagiert, also sich genau anders verhält, als es bei den KQ-Schätzern zu beobachten ist. Zudem ist es das einzige Verfahren, welches auch unter den negativen hypergeometrischen Verteilungen unter keiner Bedingung die nominale Überdeckungswahrscheinlichkeit deutlich unterschreitet, wie es dagegen bei allen anderen Verfahren festzustellen ist. Unter den symmetrischen Verteilungen mit schwachen Tails (Normal- und Gleichverteilung) zeigen die Verfahren ähnlich gute Überdeckungen, wobei die Abweichungen zum nominalen Niveau beim Perzentil- t -Verfahren in beide Richtungen etwas stärker ausfallen als dies für die anderen Bootstrap-Verfahren der Fall ist. Da das Perzentil- t -Verfahren zudem auch unter den negativen hypergeometrischen Verteilungen zu größeren Abweichungen vom nominalen Niveau führt, wird auf dieses Verfahren auch hier im folgenden nicht weiter eingegangen.

Beim NC-Verfahren sind vor allem bei den g - h -Verteilungen mit heterogenen Varianzen größere Abweichungen von der nominalen Überdeckungswahrscheinlichkeit zu beobachten. Tendenziell zeigt sich hier, daß bei den Verteilungen mit starken Tails (g - h -1 und g - h -3) der zu schätzende Parameter mit steigender Effektgröße seltener überdeckt wird, bei den schiefen Verteilungen mit weniger starken Tails hingegen häufiger überdeckt wird. Im Vergleich zwischen Perzentil- und BCA-Verfahren zeigt Abbildung 17, daß das BCA-Verfahren unter den g - h -Verteilungen ein insgesamt gleichmäßigeres Überdeckungsverhalten besitzt als das Perzentil-Verfahren und nur im

Falle der Gegenläufigkeit von Heterogenität und Disbalance die nominale Wahrscheinlichkeit deutlicher unterschreitet. Für den in Tabelle 21 nach Verteilung und Stichprobengröße aufgeteilten Vergleich ist in diesem Zusammenhang interessant, ob die höhere Überdeckungswahrscheinlichkeit des Perzentil-Verfahrens auch mit längeren Intervallen einhergeht. Empirisch besteht nämlich kein unmittelbarer Zusammenhang zwischen Überdeckungswahrscheinlichkeit und Intervalllänge (vgl. Sievers, 1998, S. 26).

Tabelle 21. Simulierte Überdeckungswahrscheinlichkeiten und Intervalllängen ausgewählter Schätzer für d'_R getrennt nach Verteilungen

			Überdeckung				Länge			
			$pCP_{.936}^{.966}$	$pCP_{.936}^{1.000}$	$ \overline{CP} - .95 $	Mittel	Min	Max	Mittel	Max
Normal	40	NC($d'_{s,R}$)	.960	.960	.006	.945	.932	.955	1.718	2.881
		Perc($d'_{s,R}$)	.920	.960	.007	.952	.935	.969	2.128	4.215
		Perc(\hat{d}'_R)	.920	1.000	.008	.953	.936	.968	1.982	3.918
		BCA($d'_{s,R}$)	.800	.800	.010	.940	.927	.950	2.137	4.102
		BCA(\hat{d}'_R)	.680	.680	.011	.939	.927	.950	1.993	3.810
	120	NC($d'_{s,R}$)	1.000	1.000	.005	.947	.936	.960	0.949	1.514
		Perc($d'_{s,R}$)	1.000	1.000	.005	.954	.944	.961	1.025	1.762
		Perc(\hat{d}'_R)	1.000	1.000	.005	.954	.943	.962	1.007	1.742
		BCA($d'_{s,R}$)	1.000	1.000	.004	.947	.939	.952	1.026	1.760
		BCA(\hat{d}'_R)	1.000	1.000	.004	.947	.939	.951	1.008	1.740
Gleich	40	NC($d'_{s,R}$)	.840	.840	.007	.943	.926	.954	2.062	3.259
		Perc($d'_{s,R}$)	1.000	1.000	.005	.946	.937	.954	2.589	4.721
		Perc(\hat{d}'_R)	1.000	1.000	.005	.948	.936	.960	2.406	4.383
		BCA($d'_{s,R}$)	.880	.880	.007	.943	.933	.953	2.555	4.457
		BCA(\hat{d}'_R)	.800	.800	.009	.941	.922	.953	2.379	4.149
	120	NC($d'_{s,R}$)	1.000	1.000	.004	.950	.939	.962	1.143	1.724
		Perc($d'_{s,R}$)	1.000	1.000	.003	.949	.942	.958	1.214	1.899
		Perc(\hat{d}'_R)	1.000	1.000	.003	.950	.942	.960	1.193	1.878
		BCA($d'_{s,R}$)	1.000	1.000	.004	.947	.940	.953	1.207	1.863
		BCA(\hat{d}'_R)	1.000	1.000	.004	.947	.940	.954	1.186	1.842
g-h-1	40	NC($d'_{s,R}$)	.600	.600	.016	.934	.892	.953	1.310	2.499
		Perc($d'_{s,R}$)	.480	.960	.015	.963	.936	.979	1.699	3.893
		Perc(\hat{d}'_R)	.560	.960	.012	.960	.936	.978	1.583	3.607
		BCA($d'_{s,R}$)	.640	.640	.011	.939	.923	.954	1.744	3.917
		BCA(\hat{d}'_R)	.680	.680	.013	.937	.921	.949	1.620	3.570
	120	NC($d'_{s,R}$)	.520	.520	.017	.934	.901	.954	0.719	1.295
		Perc($d'_{s,R}$)	.760	1.000	.010	.959	.945	.971	0.824	1.669
		Perc(\hat{d}'_R)	.840	1.000	.009	.958	.945	.971	0.809	1.651
		BCA($d'_{s,R}$)	1.000	1.000	.005	.945	.937	.954	0.831	1.692
		BCA(\hat{d}'_R)	1.000	1.000	.005	.946	.937	.954	0.817	1.672

(Fortsetzung Tabelle 21)

			Überdeckung					Länge		
			$pCP_{.936}^{.966}$	$pCP_{.936}^{1.000}$	$ CP - .95 $	Mittel	Min	Max	Mittel	Max
g-h-2	40	NC($d'_{s,R}$)	.360	.760	.020	.952	.900	.982	1.454	2.454
		Perc($d'_{s,R}$)	.560	1.000	.013	.961	.937	.982	1.740	3.201
		Perc(\hat{d}'_R)	.600	1.000	.014	.962	.937	.988	1.620	2.975
		BCA($d'_{s,R}$)	.800	.800	.009	.941	.917	.952	1.759	3.151
		BCA(\hat{d}'_R)	.640	.640	.011	.939	.918	.951	1.640	2.913
	120	NC($d'_{s,R}$)	.320	.760	.024	.957	.892	.990	0.802	1.291
		Perc($d'_{s,R}$)	.680	1.000	.012	.960	.943	.975	0.815	1.253
		Perc(\hat{d}'_R)	.720	1.000	.012	.960	.941	.978	0.801	1.239
		BCA($d'_{s,R}$)	.960	.960	.005	.946	.933	.955	0.818	1.261
		BCA(\hat{d}'_R)	.960	.960	.005	.946	.933	.953	0.804	1.246
g-h-3	40	NC($d'_{s,R}$)	.800	.800	.009	.944	.911	.961	1.236	2.354
		Perc($d'_{s,R}$)	.440	1.000	.018	.966	.937	.987	1.575	3.569
		Perc(\hat{d}'_R)	.560	1.000	.016	.964	.937	.984	1.467	3.310
		BCA($d'_{s,R}$)	.720	.720	.009	.941	.921	.956	1.622	3.609
		BCA(\hat{d}'_R)	.640	.640	.012	.939	.919	.953	1.508	3.291
	120	NC($d'_{s,R}$)	.800	.800	.012	.945	.901	.962	0.679	1.229
		Perc($d'_{s,R}$)	.600	1.000	.012	.962	.946	.979	0.753	1.507
		Perc(\hat{d}'_R)	.720	1.000	.011	.961	.946	.980	0.740	1.491
		BCA($d'_{s,R}$)	.960	.960	.005	.945	.933	.955	0.761	1.532
		BCA(\hat{d}'_R)	.960	.960	.005	.946	.936	.953	0.747	1.515
g-h-4	40	NC($d'_{s,R}$)	.480	.800	.020	.956	.892	.988	1.792	2.777
		Perc($d'_{s,R}$)	.840	.920	.006	.951	.932	.971	2.103	3.453
		Perc(\hat{d}'_R)	.720	.920	.010	.955	.932	.983	1.957	3.202
		BCA($d'_{s,R}$)	.840	.840	.009	.941	.918	.954	2.087	3.488
		BCA(\hat{d}'_R)	.760	.760	.010	.941	.922	.954	1.945	3.240
	120	NC($d'_{s,R}$)	.360	.840	.023	.963	.917	.993	0.994	1.488
		Perc($d'_{s,R}$)	.920	1.000	.007	.955	.944	.971	0.970	1.496
		Perc(\hat{d}'_R)	.880	1.000	.008	.957	.946	.977	0.953	1.477
		BCA($d'_{s,R}$)	.960	.960	.005	.946	.933	.954	0.968	1.493
		BCA(\hat{d}'_R)	.920	.920	.005	.946	.934	.954	0.951	1.474
Hyper	40	NC($d'_{s,R}$)	.640	.640	.020	.933	.827	.963	2.143	3.341
		Perc($d'_{s,R}$)	.800	.960	.009	.953	.931	.980	3.037	6.035
		Perc(\hat{d}'_R)	.840	.920	.009	.949	.909	.976	2.842	5.687
		BCA($d'_{s,R}$)	.520	.520	.024	.927	.833	.958	2.930	5.758
		BCA(\hat{d}'_R)	.440	.440	.029	.922	.833	.957	2.742	5.436
	120	NC($d'_{s,R}$)	.720	.720	.019	.934	.795	.957	1.176	1.758
		Perc($d'_{s,R}$)	.960	1.000	.005	.950	.941	.970	1.352	2.277
		Perc(\hat{d}'_R)	.880	.920	.006	.948	.909	.969	1.327	2.245
		BCA($d'_{s,R}$)	.640	.640	.016	.936	.892	.959	1.339	2.195
		BCA(\hat{d}'_R)	.520	.520	.019	.933	.853	.959	1.314	2.165

Da das NC-Verfahren bei den hier simulierten Modellen nur unter der Normal- und der Gleichverteilung zu ähnlich guten Überdeckungsrate führt wie die Bootstrap-Verfahren, kommt dieses als Kandidat für eine generelle Empfehlung zur Schätzung von d'_R trotz der bei kleineren Stichproben

deutlich kürzeren Intervalle nicht in Frage. Eine Entscheidung zwischen den Bootstrap-Verfahren erscheint hier einfach möglich, da die Perzentilverfahren trotz konservativeren Verhaltens bei $N = 40$ durchgängig kürzere Intervalle aufweisen als die BCA-Verfahren. Zudem halten die Perzentil-Verfahren als einzige die nominale Wahrscheinlichkeit auch bei den diskreten Verteilungen ein. Da zudem keine wesentlichen Unterschiede hinsichtlich der Überdeckung zwischen den beiden Perzentil-Verfahren auftreten, kann für die hier realisierten Bedingungen das Perzentil-Verfahren mit korrigiertem Punktschätzer als generelle Empfehlung für die Intervallschätzung von d'_R gewählt werden.

2.1.5.2.4 Schätzer für $d_R^{S_1}$ und $d_R^{S_2}$

Klarer als bei den bisher untersuchten Schätzern macht hier die Gesamtauswertung (Tabelle 22) zu den Schätzverfahren von $d_R^{S_1}$ und $d_R^{S_2}$ bereits deutlich, daß hinsichtlich der Überdeckungsleistung nur die Perzentil- oder BCA-Verfahren als geeignete Schätzer in Frage kommen, denn insbesondere bei der Schätzung von $d_R^{S_2}$ liegen das NC-Verfahren sowie die Perzentil- t -Verfahren erheblich unterhalb der nominalen Überdeckungswahrscheinlichkeit. Die untere Grenze des Robustheitskriteriums .936 wird in jeweils etwa 60% aller Simulationsbedingungen unterschritten (vgl. $pCP_{.936}^{1.000}$ in Tabelle 22) und die mittleren absoluten Fehler ($|\overline{CP} - .95|$) liegen zwischen .034 und .038. Betrachtet man die minimale und maximale Überdeckungswahrscheinlichkeit, wird auch hier das konservativere Verhalten der Perzentil-Verfahren im Vergleich zu den BCA-Verfahren offensichtlich. Die mittleren und insbesondere maximalen Längen der Bootstrap-Intervalle weisen jedoch auf die für das Bootstrap bereits bei der Untersuchung zur Punktschätzung deutlich gewordene Anforderung einer hinreichend großen Stichprobe hin, die in Verbindung mit 20%-robustifizierten Statistiken und einem $n = 10$ in der kleineren Gruppe hier nicht erfüllt ist.

Tabelle 22. Simulierte Überdeckungswahrscheinlichkeiten und Intervalllängen der Schätzer für $d_R^{S_1}$ und $d_R^{S_2}$ über alle 350 Simulationsbedingungen

	Überdeckung						Länge	
	$pCP_{.936}^{.966}$	$pCP_{.936}^{1.000}$	$ \overline{CP} - .95 $	Mittel	Min	Max	Mittel	Max
NC($d_{s,R}^{S_1}$)	.594	.806	.016	.952	.892	.994	1.427	3.566
Perc($d_{s,R}^{S_1}$)	.731	<u>.986</u>	.010	.957	.931	.984	2.403	11.142
BCA($d_{s,R}^{S_1}$)	<u>.914</u>	.914	<u>.007</u>	.945	.837	.962	1.916	11.891
Perc-t($d_{s,R}^{S_1}$)	.474	.723	.020	.948	.824	.999	1.648	5.527
Perc($\hat{d}_R^{S_1}$)	.649	<u>.986</u>	.013	.961	.925	.992	2.154	9.383
BCA($\hat{d}_R^{S_1}$)	.909	.909	<u>.007</u>	.946	.836	.963	1.747	8.486
Perc-t($\hat{d}_R^{S_1}$)	.374	.586	.028	.937	.775	.999	1.501	4.654
NC($d_{s,R}^{S_2}$)	.377	.417	.038	.916	.450	.988	1.396	3.343
Perc($d_{s,R}^{S_2}$)	<u>.851</u>	.986	<u>.008</u>	.955	.930	.980	2.150	15.344
BCA($d_{s,R}^{S_2}$)	.846	.849	.010	.943	.807	.969	1.930	16.159
Perc-t($d_{s,R}^{S_2}$)	.357	.406	.034	.920	.478	.993	1.845	7.531
Perc($\hat{d}_R^{S_2}$)	.803	<u>.989</u>	.010	.957	.931	.993	1.989	12.921
BCA($\hat{d}_R^{S_2}$)	.846	.857	.010	.944	.802	.971	1.819	14.053
Perc-t($\hat{d}_R^{S_2}$)	.331	.374	.038	.915	.465	.994	1.726	7.193

Anmerkung: Unterstreichungen kennzeichnen die besten Resultate je Kriterium

Einen genaueren Überblick zum Überdeckungsverhalten der einzelnen Verfahren bei $N = 40$ geben nun Abbildung 18 und Abbildung 19.

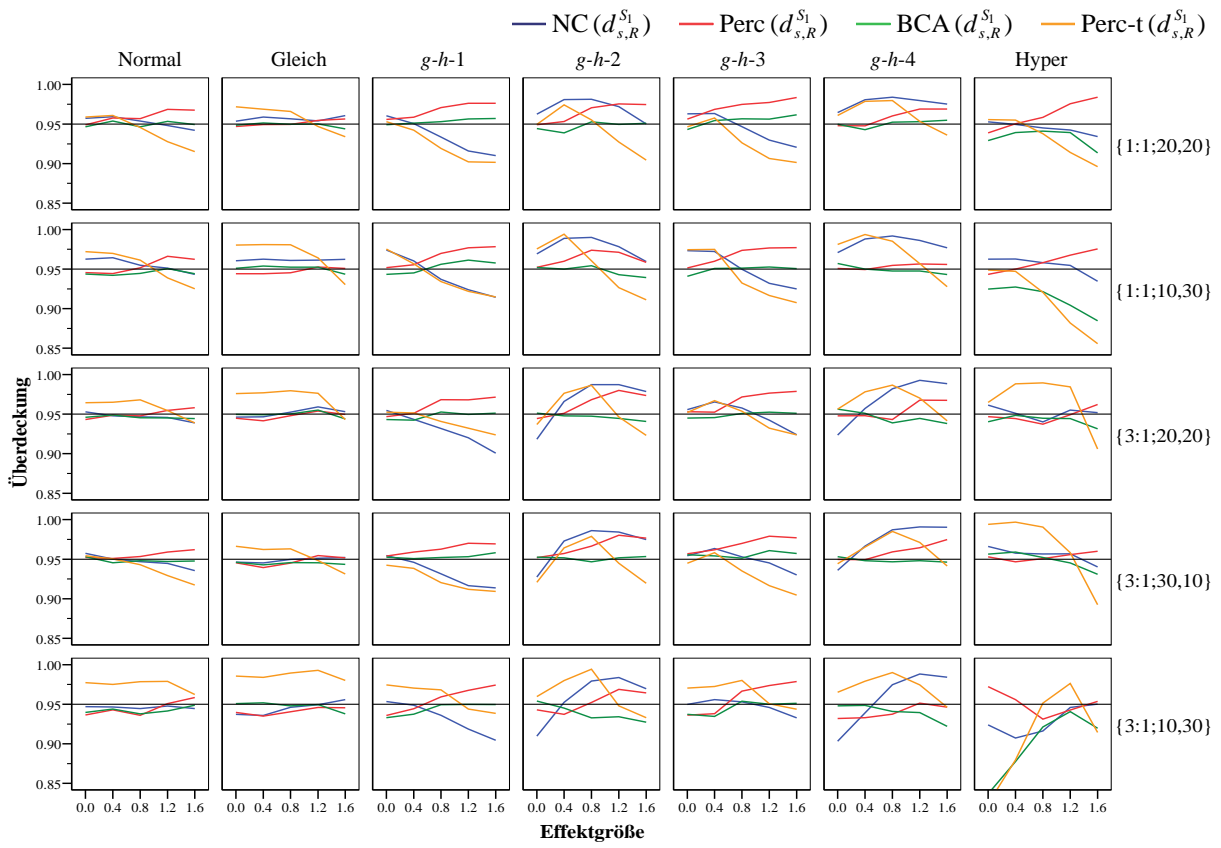


Abbildung 18. Simulierte Überdeckungswahrscheinlichkeiten ausgewählter Schätzer für d_R^{S1}

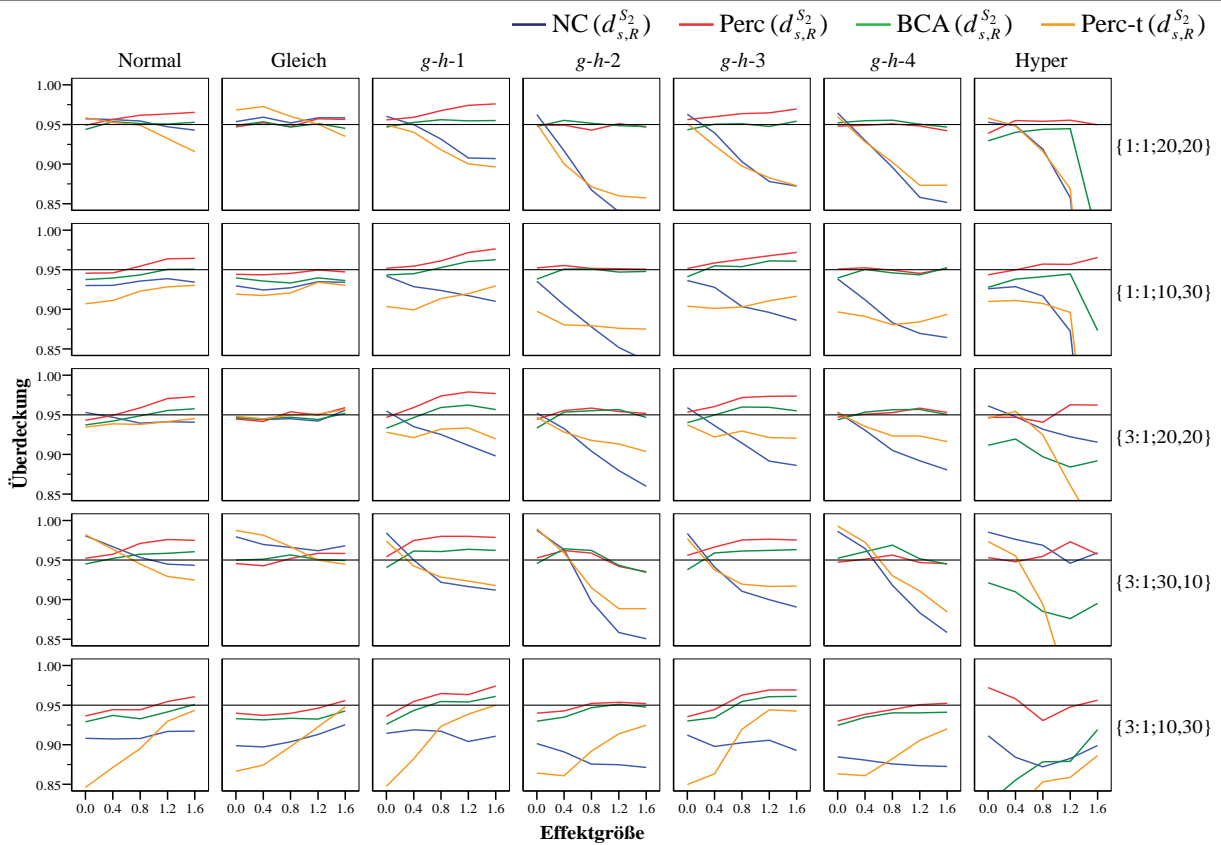


Abbildung 19. Simulierte Überdeckungswahrscheinlichkeiten ausgewählter Schätzer für d_R^{S2}

Es sei daran erinnert, daß nur unter den Bedingungen mit symmetrischen Verteilungen, gleich großen Gruppen und homogenen Varianzen (vgl. jeweils obere Reihe Normal, Gleich und $g-h-1$ in Abbildung 18 und Abbildung 19) die Schätzer für $d_R^{S_1}$ und $d_R^{S_2}$ sowohl denselben Parameter als auch denselben Standardfehler aufweisen. Das Perzentil-Verfahren ist hier das einzige, welches die nominale Überdeckungswahrscheinlichkeit für die Schätzung von $d_R^{S_1}$ und $d_R^{S_2}$ unter keiner der simulierten Bedingungen bei $N = 40$ drastisch unterschreitet. Das BCA-Verfahren liefert bei den stetigen Verteilungen eine mit dem Perzentil-Verfahren vergleichbar gute Überdeckungsleistung, ist jedoch zur Schätzung unter der diskreten negativen hypergeometrischen Verteilung nicht geeignet. Wie die nach Verteilungen getrennte Auswertung in Tabelle 23 und Tabelle 24 aufzeigt, bietet sich von den hier ausgewählten Verfahren das BCA-Bootstrap zum Einsatz unter stetigen Verteilungen an. Bei kleineren Stichproben führt das ähnlich gut überdeckende Perzentil-Bootstrap zu zum Teil deutlich längeren Intervallen. Unter der hier simulierten Familie diskreter Verteilungen hingegen ist das Perzentil-Bootstrap das einzige Verfahren, das zu akzeptablen Überdeckungswahrscheinlichkeiten führt.

Tabelle 23. Simulierte Überdeckungswahrscheinlichkeiten und Intervallängen ausgewählter Schätzer für $d_R^{S_1}$ getrennt nach Verteilungen

			Überdeckung					Länge		
			$pCP_{.936}^{.966}$	$pCP_{.936}^{1.000}$	$ CP - .95 $	Mittel	Min	Max	Mittel	Max
Normal	40	Perc($d_{s,R}^{S_1}$)	.880	1.000	.007	.953	.936	.969	3.712	9.261
		Perc($\hat{d}_R^{S_1}$)	.760	1.000	.012	.959	.936	.981	3.245	7.798
		BCA($d_{s,R}^{S_1}$)	1.000	1.000	.005	.946	.938	.954	2.896	7.298
		BCA($\hat{d}_R^{S_1}$)	.960	.960	.004	.948	.935	.957	2.627	6.301
	120	Perc($d_{s,R}^{S_1}$)	.960	1.000	.006	.955	.944	.967	1.172	2.042
		Perc($\hat{d}_R^{S_1}$)	.920	1.000	.007	.956	.943	.970	1.134	1.950
		BCA($d_{s,R}^{S_1}$)	1.000	1.000	.003	.948	.941	.953	1.173	2.045
		BCA($\hat{d}_R^{S_1}$)	1.000	1.000	.003	.948	.941	.956	1.136	1.958
Gleich	40	Perc($d_{s,R}^{S_1}$)	.960	.960	.005	.947	.935	.956	4.500	10.284
		Perc($\hat{d}_R^{S_1}$)	.840	1.000	.009	.954	.938	.976	3.928	8.661
		BCA($d_{s,R}^{S_1}$)	1.000	1.000	.003	.948	.938	.955	3.856	8.879
		BCA($\hat{d}_R^{S_1}$)	1.000	1.000	.004	.949	.942	.957	3.513	7.809
	120	Perc($d_{s,R}^{S_1}$)	1.000	1.000	.003	.950	.942	.959	1.346	2.135
		Perc($\hat{d}_R^{S_1}$)	1.000	1.000	.004	.951	.942	.963	1.302	2.039
		BCA($d_{s,R}^{S_1}$)	1.000	1.000	.004	.947	.940	.954	1.331	2.087
		BCA($\hat{d}_R^{S_1}$)	1.000	1.000	.004	.947	.938	.955	1.289	1.998
$g-h-1$	40	Perc($d_{s,R}^{S_1}$)	.440	.960	.015	.963	.936	.978	3.219	9.281
		Perc($\hat{d}_R^{S_1}$)	.400	.960	.019	.967	.936	.992	2.812	7.815
		BCA($d_{s,R}^{S_1}$)	.960	.960	.005	.950	.933	.961	1.839	3.116
		BCA($\hat{d}_R^{S_1}$)	.960	.960	.006	.951	.927	.963	1.727	2.907
	120	Perc($d_{s,R}^{S_1}$)	.720	1.000	.010	.959	.945	.972	0.995	2.002
		Perc($\hat{d}_R^{S_1}$)	.680	1.000	.011	.960	.945	.973	0.962	1.913
		BCA($d_{s,R}^{S_1}$)	1.000	1.000	.003	.948	.940	.954	1.010	2.071
		BCA($\hat{d}_R^{S_1}$)	1.000	1.000	.003	.949	.942	.956	0.978	1.983

(Fortsetzung Tabelle 23)

			Überdeckung						Länge	
			$pCP_{.936}^{.966}$	$pCP_{.936}^{1.000}$	$ CP-.95 $	Mittel	Min	Max	Mittel	Max
<i>g-h-2</i>	40	Perc($d_{s,R}^{S_1}$)	.520	1.000	.014	.962	.937	.980	3.183	9.729
		Perc($\hat{d}_{R}^{S_1}$)	.360	1.000	.022	.970	.937	.991	2.778	8.193
		BCA($d_{s,R}^{S_1}$)	.880	.880	.006	.946	.927	.954	2.173	5.971
		BCA($\hat{d}_{R}^{S_1}$)	.960	.960	.005	.948	.930	.959	1.996	5.215
	120	Perc($d_{s,R}^{S_1}$)	.560	1.000	.013	.962	.943	.976	0.919	1.861
		Perc($\hat{d}_{R}^{S_1}$)	.520	1.000	.016	.964	.945	.983	0.889	1.777
		BCA($d_{s,R}^{S_1}$)	.960	.960	.005	.947	.934	.955	0.917	1.828
		BCA($\hat{d}_{R}^{S_1}$)	1.000	1.000	.004	.947	.938	.955	0.888	1.749
<i>g-h-3</i>	40	Perc($d_{s,R}^{S_1}$)	.360	1.000	.019	.967	.936	.984	2.966	9.062
		Perc($\hat{d}_{R}^{S_1}$)	.320	1.000	.022	.971	.936	.992	2.590	7.631
		BCA($d_{s,R}^{S_1}$)	.960	.960	.005	.951	.935	.962	1.705	3.051
		BCA($\hat{d}_{R}^{S_1}$)	.920	.920	.006	.952	.934	.961	1.602	2.853
	120	Perc($d_{s,R}^{S_1}$)	.600	1.000	.013	.963	.947	.980	0.902	1.909
		Perc($\hat{d}_{R}^{S_1}$)	.560	1.000	.014	.964	.947	.981	0.873	1.824
		BCA($d_{s,R}^{S_1}$)	1.000	1.000	.004	.948	.940	.960	0.919	1.986
		BCA($\hat{d}_{R}^{S_1}$)	1.000	1.000	.004	.948	.942	.957	0.890	1.902
<i>g-h-4</i>	40	Perc($d_{s,R}^{S_1}$)	.720	.920	.009	.953	.932	.975	4.092	11.142
		Perc($\hat{d}_{R}^{S_1}$)	.440	.920	.017	.964	.933	.989	3.561	9.383
		BCA($d_{s,R}^{S_1}$)	.960	.960	.005	.947	.922	.957	3.549	11.891
		BCA($\hat{d}_{R}^{S_1}$)	.960	.960	.005	.949	.932	.963	2.737	8.486
	120	Perc($d_{s,R}^{S_1}$)	.840	1.000	.009	.958	.944	.971	1.057	1.831
		Perc($\hat{d}_{R}^{S_1}$)	.600	1.000	.013	.962	.948	.982	1.022	1.749
		BCA($d_{s,R}^{S_1}$)	.960	.960	.005	.947	.933	.954	1.047	1.735
		BCA($\hat{d}_{R}^{S_1}$)	.960	.960	.004	.948	.932	.955	1.014	1.661
Hyper	40	Perc($d_{s,R}^{S_1}$)	.760	.960	.010	.955	.931	.984	4.097	8.973
		Perc($\hat{d}_{R}^{S_1}$)	.800	1.000	.011	.956	.936	.986	3.619	7.557
		BCA($d_{s,R}^{S_1}$)	.480	.480	.024	.927	.837	.959	3.846	8.331
		BCA($\hat{d}_{R}^{S_1}$)	.440	.440	.027	.924	.836	.958	3.429	7.119
	120	Perc($d_{s,R}^{S_1}$)	.920	1.000	.008	.952	.937	.976	1.482	2.806
		Perc($\hat{d}_{R}^{S_1}$)	.880	.920	.007	.950	.925	.974	1.433	2.680
		BCA($d_{s,R}^{S_1}$)	.640	.640	.020	.932	.868	.959	1.461	2.630
		BCA($\hat{d}_{R}^{S_1}$)	.560	.560	.023	.929	.861	.959	1.414	2.516

Tabelle 24. Simulierte Überdeckungswahrscheinlichkeiten und Intervallängen ausgewählter Schätzer für $d_R^{S_2}$ getrennt nach Verteilungen

			Überdeckung						Länge	
			$pCP_{.936}^{.966}$	$pCP_{.936}^{1.000}$	$ CP-.95 $	Mittel	Min	Max	Mittel	Max
Normal	40	Perc($d_{s,R}^{S_2}$)	.800	1.000	.011	.957	.936	.976	2.920	8.964
		Perc($\hat{d}_R^{S_2}$)	.720	1.000	.014	.961	.936	.990	2.642	7.549
		BCA($d_{s,R}^{S_2}$)	.920	.920	.007	.947	.929	.961	2.312	3.147
		BCA($\hat{d}_R^{S_2}$)	.920	.920	.008	.949	.929	.965	2.179	2.931
	120	Perc($d_{s,R}^{S_2}$)	.880	1.000	.006	.955	.944	.971	1.107	1.867
		Perc($\hat{d}_R^{S_2}$)	.880	1.000	.007	.956	.942	.972	1.081	1.784
		BCA($d_{s,R}^{S_2}$)	1.000	1.000	.003	.948	.941	.953	1.109	1.882
		BCA($\hat{d}_R^{S_2}$)	1.000	1.000	.003	.948	.941	.953	1.083	1.803
Gleich	40	Perc($d_{s,R}^{S_2}$)	1.000	1.000	.006	.949	.937	.959	3.311	10.177
		Perc($\hat{d}_R^{S_2}$)	.800	1.000	.010	.954	.937	.984	2.992	8.570
		BCA($d_{s,R}^{S_2}$)	.760	.760	.007	.944	.931	.956	2.657	5.540
		BCA($\hat{d}_R^{S_2}$)	.720	.760	.008	.946	.932	.968	2.480	4.715
	120	Perc($d_{s,R}^{S_2}$)	1.000	1.000	.003	.951	.942	.962	1.252	1.815
		Perc($\hat{d}_R^{S_2}$)	.960	1.000	.004	.952	.942	.971	1.222	1.734
		BCA($d_{s,R}^{S_2}$)	1.000	1.000	.003	.948	.941	.952	1.240	1.741
		BCA($\hat{d}_R^{S_2}$)	1.000	1.000	.003	.948	.945	.953	1.212	1.666
g-h-1	40	Perc($d_{s,R}^{S_2}$)	.440	.960	.017	.966	.936	.980	2.594	8.918
		Perc($\hat{d}_R^{S_2}$)	.440	.960	.019	.967	.936	.993	2.343	7.510
		BCA($d_{s,R}^{S_2}$)	.920	.920	.009	.953	.926	.964	2.043	3.234
		BCA($\hat{d}_R^{S_2}$)	.880	.920	.009	.953	.926	.966	1.926	3.017
	120	Perc($d_{s,R}^{S_2}$)	.760	1.000	.011	.961	.945	.976	0.934	1.935
		Perc($\hat{d}_R^{S_2}$)	.760	1.000	.011	.960	.945	.978	0.911	1.848
		BCA($d_{s,R}^{S_2}$)	.960	.960	.004	.949	.936	.957	0.945	2.002
		BCA($\hat{d}_R^{S_2}$)	.960	.960	.004	.950	.936	.956	0.922	1.917
g-h-2	40	Perc($d_{s,R}^{S_2}$)	.960	.960	.005	.950	.935	.962	3.275	10.978
		Perc($\hat{d}_R^{S_2}$)	.960	1.000	.008	.956	.940	.969	2.960	9.245
		BCA($d_{s,R}^{S_2}$)	.840	.840	.007	.948	.930	.964	3.436	14.123
		BCA($\hat{d}_R^{S_2}$)	.960	.960	.006	.949	.931	.966	3.130	11.937
	120	Perc($d_{s,R}^{S_2}$)	1.000	1.000	.004	.952	.946	.958	1.158	2.199
		Perc($\hat{d}_R^{S_2}$)	1.000	1.000	.005	.954	.946	.961	1.130	2.101
		BCA($d_{s,R}^{S_2}$)	1.000	1.000	.003	.949	.942	.956	1.138	2.112
		BCA($\hat{d}_R^{S_2}$)	1.000	1.000	.003	.950	.942	.959	1.111	2.022
g-h-3	40	Perc($d_{s,R}^{S_2}$)	.480	.960	.015	.964	.935	.976	2.552	8.885
		Perc($\hat{d}_R^{S_2}$)	.440	1.000	.018	.967	.937	.990	2.307	7.482
		BCA($d_{s,R}^{S_2}$)	.920	.920	.008	.952	.930	.963	2.068	3.287
		BCA($\hat{d}_R^{S_2}$)	.880	.920	.009	.953	.930	.968	1.936	3.067
	120	Perc($d_{s,R}^{S_2}$)	.800	1.000	.009	.959	.946	.970	0.924	1.917
		Perc($\hat{d}_R^{S_2}$)	.800	1.000	.010	.960	.945	.973	0.901	1.831
		BCA($d_{s,R}^{S_2}$)	1.000	1.000	.004	.949	.938	.959	0.931	1.964
		BCA($\hat{d}_R^{S_2}$)	1.000	1.000	.004	.949	.938	.959	0.908	1.880

(Fortsetzung Tabelle 24)

		Überdeckung						Länge		
		$pCP_{.936}^{.966}$	$pCP_{.936}^{1.000}$	$ CP-.95 $	Mittel	Min	Max	Mittel	Max	
g-h-4	40	Perc($d_{s,R}^{S_2}$)	.960	.960	.004	.948	.930	.958	4.301	15.344
		Perc($\hat{d}_R^{S_2}$)	.920	.960	.006	.953	.931	.967	3.868	12.921
		BCA($d_{s,R}^{S_2}$)	.880	.920	.007	.948	.925	.969	4.103	16.159
		BCA($\hat{d}_R^{S_2}$)	.880	.920	.007	.949	.925	.971	3.755	14.053
	120	Perc($d_{s,R}^{S_2}$)	1.000	1.000	.003	.952	.940	.960	1.371	2.442
		Perc($\hat{d}_R^{S_2}$)	1.000	1.000	.004	.952	.940	.960	1.338	2.333
		BCA($d_{s,R}^{S_2}$)	1.000	1.000	.003	.950	.939	.957	1.342	2.306
		BCA($\hat{d}_R^{S_2}$)	1.000	1.000	.003	.950	.939	.957	1.310	2.207
Hyper	40	Perc($d_{s,R}^{S_2}$)	.880	.960	.008	.953	.931	.973	2.764	3.878
		Perc($\hat{d}_R^{S_2}$)	.720	.920	.010	.955	.933	.977	2.552	3.704
		BCA($d_{s,R}^{S_2}$)	.240	.240	.048	.902	.807	.945	2.646	3.668
		BCA($\hat{d}_R^{S_2}$)	.240	.240	.045	.905	.802	.946	2.442	3.415
	120	Perc($d_{s,R}^{S_2}$)	.960	1.000	.008	.953	.937	.971	1.636	3.211
		Perc($\hat{d}_R^{S_2}$)	.840	1.000	.008	.954	.941	.976	1.598	3.142
		BCA($d_{s,R}^{S_2}$)	.400	.400	.029	.922	.845	.958	1.585	3.087
		BCA($\hat{d}_R^{S_2}$)	.400	.400	.028	.924	.855	.958	1.549	3.021

2.1.5.2.5 Vergleich von KQ- und robustifizierter Schätzung

Abschließend seien die Eigenschaften der jeweils am besten geeigneten Intervallschätzer der einzelnen Parameter gegenübergestellt. Abbildung 20a zeigt die Verteilungen der Überdeckungswahrscheinlichkeiten von $BCA(\hat{d}')$ und $Perc(\hat{d}'_R)$. Während die Schätzungen von d' und d'_R bei Verteilungen mit Tailstärken nahe derjenigen der Normalverteilung (Gleich, g-h-4 und Hyper) etwa gleich gut gelingen, kann mit den hier simulierten Verfahren kein robustes Intervall für d' bei Verteilungen mit stärkeren Tails (g-h-1, g-h-2 und g-h-3) konstruiert werden. Zu beachten ist, daß die Verteilung g-h-2 hinsichtlich Schiefe und Tailstärke der Exponentialverteilung entspricht und also keine extremen, sondern z.B. für Reaktionszeiten praxisrelevante Charakteristika aufweist.

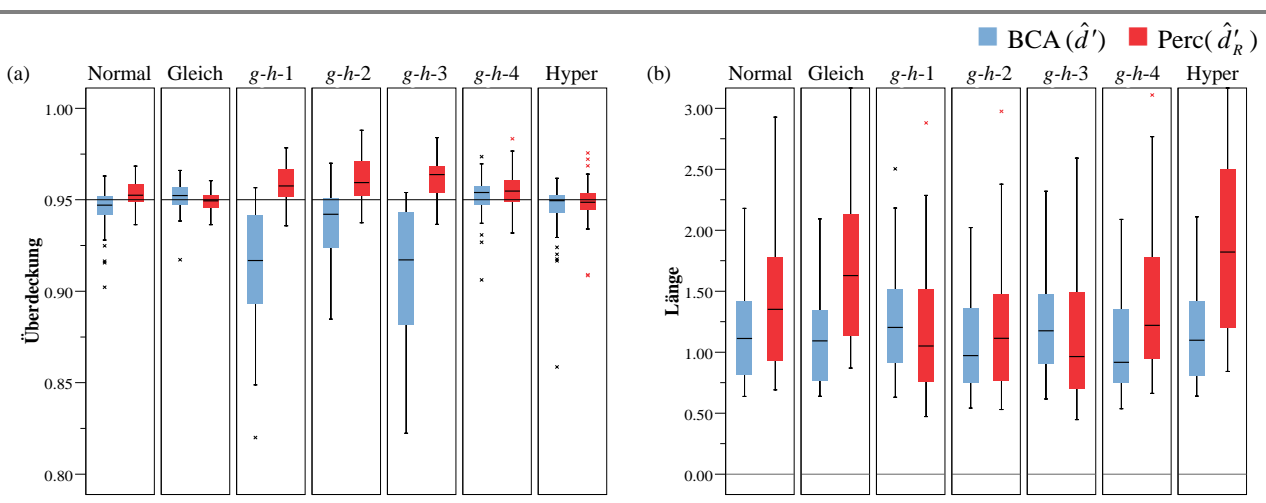


Abbildung 20. Simulierte Überdeckungswahrscheinlichkeiten und Intervalllängen von $BCA(\hat{d}')$ und $Perc(\hat{d}'_R)$ getrennt nach Verteilungen (die Balkenenden der Plots geben jeweils die größten und kleinsten beobachteten Werte an, die keine Ausreißer nach Tukey (1977) [1.5-facher Interquartil-Abstand vom jeweiligen Quartil] sind)

In Abbildung 20b sind die Verteilungen der simulierten Intervalllängen der beiden Verfahren gegenübergestellt. Hier wird deutlich, daß die Intervalle auf Basis robustifizierter Statistiken bei Verteilungen mit schwachen Tails und insbesondere unter dem diskreten Modell deutlich länger und damit weniger präzise sind als die Intervalle auf Basis der KQ-Schätzer, die unter der Normalverteilung optimale Eigenschaften besitzen. Bei den folgenden Gegenüberstellungen (Abbildung 21) wurden die Ergebnisse von $\text{BCA}(\hat{d}^{S_1})$ und $\text{BCA}(\hat{d}^{S_2})$ sowie von $\text{Perc}(\hat{d}_R^{S_1})$ und $\text{Perc}(\hat{d}_R^{S_2})$ jeweils zusammengefaßt.

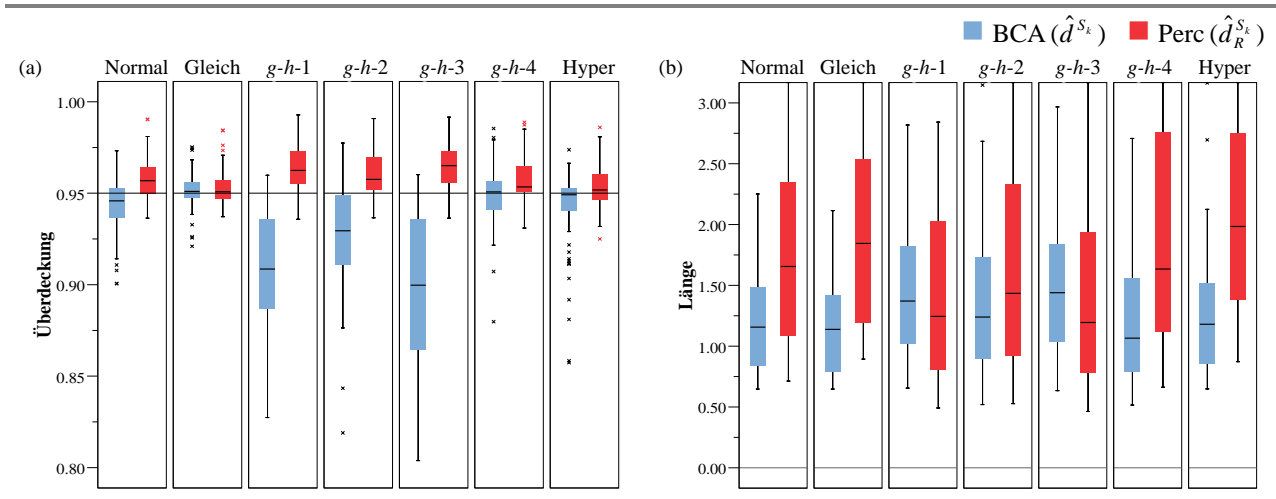


Abbildung 21. Simulierte Überdeckungswahrscheinlichkeiten und Intervalllängen von $\text{BCA}(\hat{d}^{S_k})$ und $\text{Perc}(\hat{d}_R^{S_k})$ getrennt nach Verteilungen

Dies erscheint hier sinnvoll, da die Simulationsbedingungen bei beiden Schätzertypen dieselben sind. Jedem Boxplot liegen also 100 Simulationsergebnisse zugrunde. Die Ergebnisstruktur zur Überdeckung gleicht insgesamt derjenigen aus Abbildung 20 zu den Schätzern für d' und d'_R . In Abbildung 21b wird jedoch die Schwäche des Perzentil-Bootstraps bei zu kleinen Stichproben ersichtlich, die sich hier in deutlich längeren Intervallen ausdrückt.

2.1.5.3 Fazit

Die Ergebnisse der dargestellten Simulationsstudie zu standardisierten Lagedifferenzen bei unabhängigen Stichproben haben deutlich gemacht, dass die Einhaltung des Konfidenzniveaus der KQ-Schätzer ganz wesentlich von der zugrundeliegenden Verteilung und der Größe des zu schätzenden Parameterwertes sowie auch von den Varianzen (gleich/ungleich) und den Stichprobenumfängen (Gesamtumfang, balanciert/unbalanciert) beeinflusst wird. Zentraler Befund ist, daß die herkömmlichen Parameter d' und d^{S_k} anhand der hier untersuchten Verfahren unter den betrachteten Verteilungen mit stärkeren Tails ($g-h-1$, $g-h-2$ und $g-h-3$) nicht robust geschätzt werden können. Vielmehr unterschreiten sowohl die nichtzentralen als auch die Bootstrap-Schätzer das nominale Konfidenzniveau unter diesen Verteilungen selbst bei $N = 120$ zum Teil sehr deutlich. Hingegen gelingt die Schätzung der (20%-)robustifizierten Größen insbesondere mit dem Perzentil-Bootstrap über alle Verteilungen hinweg relativ robust. Anders als die KQ-Schätzer neigen diese Verfahren unter den tailstarken Verteilungen zu einem konservativen Überdeckungsverhalten, d.h. das nominale Konfidenzniveau wird tendentiell überschritten (vgl. Abbildung 20 und Abbildung 21). Zusammenfassend lassen sich dem Anwender folgende Empfehlungen machen:

- Sofern von Normalverteilungen oder von Verteilungen ausgegangen werden kann, deren Tailstärken nicht oberhalb derjenigen der Normalverteilung liegen, empfiehlt sich die Schätzung der herkömmlichen Parameter d' und d^{S_k} anhand der nichtzentralen Schätzer $\text{NC}(d'_s)$ und $\text{NC}(d^{S_2})$. Für die Punktschätzungen bieten sich die biaskorrigierten Statistiken \hat{d}' und \hat{d}^{S_k} an.
- Sofern auch von Verteilungen mit stärkeren Tails auszugehen ist, verbietet sich die Schätzung der herkömmlichen Parameter. In diesem Fall sollte einer robustifizierten Schätzung der Parameter d'_R oder $d^{S_k}_R$ anhand des Perzentil-Bootstraps der Vorzug gegeben werden. Bei kleinen Stichproben mit weniger als 20 Werten je Gruppe sollte die Punktschätzung anhand von \hat{d}'_R und $\hat{d}^{S_k}_R$ vorgenommen werden, bei größeren Stichproben bieten sich auch die Bootstrap-korrigierten Statistiken \hat{d}'^{Boot}_R und $\hat{d}^{S_k,Boot}$ an.

Die grundsätzliche Empfehlung für eine robustifizierte Auswertung, wie sie z. B. von Algina, Keselman und Penfield (2005a) ausgesprochen wird, wird aufgrund der vorliegenden Ergebnisse jedoch nicht gemacht, da die Intervalle der robustifizierten Schätzungen unter Verteilungen mit kurzen Tails im Durchschnitt etwa 1,5 bis 2 mal so breit sind wie die der KQ-Schätzungen (siehe Tabelle 16 im Vergleich mit Tabelle 21). Obwohl die Einhaltung des Konfidenzniveaus sicherlich das primäre Kriterium für die Wahl eines Intervallschätzers ist, ist ein solcher Präzisions- bzw. Effizienzverlust z. B. bei Daten, wie sie in der Psychologie nicht untypisch sind und hier mit der negativen hypergeometrischen Verteilung modelliert wurden, nicht zu rechtfertigen.

Für die konkrete Vorgehensweise in einer Datensituation, für welche kein Vorwissen der zugrundeliegenden Verteilungen gegeben ist und auch keine diesbezüglichen Vorannahmen getroffen wurden, gibt die vorgelegte Studie jedoch keine genaueren Anhaltspunkte, wie eine robuste und möglichst auch effiziente Schätzung der Effektgröße vorzunehmen ist. Von der bis heute mitunter üblichen Praxis, die Wahl einer Test- oder Schätzstatistik von den Eigenschaften der gegebenen Daten abhängig zu machen, wird abgeraten, sofern das Gesamtniveau dieser gestuften, auch adaptiv genannten Prozedur nicht bekannt und kontrollierbar ist. Eine adaptive und robuste Schätzprozedur, die das Gesamtniveau zu kontrollieren erlaubt, kann aus den vorgelegten Ergebnissen jedoch im Nachhinein nicht entwickelt werden (vgl. Kapitel 5).

2.2 Verbundene Stichproben

Neben der für den t -Test bei abhängigen Stichproben definierten Effektgröße d_D werden im folgenden auch die für unverbundene Stichproben dargestellten Effektgrößen (siehe Kapitel 2.1.1) in diesem einfachen Vortest-Nachtest-Design des t -Tests untersucht. Auf die Konstruktion und Absicherung von Effektgrößen im Vortest-Nachtest-Kontrollgruppen-Plan wird hier nicht eingegangen. Erste Ansätze und Ergebnisse zur Robustheit verschiedener Schätzer finden sich bei Morris (2008).

2.2.1 Parameter und Punktschätzer

Da die Verfahren zur Biaskorrektur und zur Intervallschätzung anhand des Bootstraps dieselben wie im unabhängigen Fall sind (vgl. Kap. 2.1.1.5 und Kap. 2.1.2.3), werden diese im folgenden nicht erneut dargestellt.

2.2.1.1 Cohens d_D

Für den t -Test zweier abhängiger Mittelwerte, der ein Spezialfall des Einstichproben- t -Tests ist, findet im Rahmen der Teststärkeanalyse als Effektgröße folgende Statistik Verwendung (Cohen, 1988, S. 48):

$$d_D = \frac{\mu_D}{\sigma_D} = \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}}.$$

Als Schätzer dient nach Cohen (S. 73) wiederum die anhand der Stichprobenwerte analog gestaltete Formel:

$$d_{D,s} = \frac{\bar{X}_D}{S_D} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_1^2 + S_2^2 - 2S_{12}}},$$

wobei S_{12} die Kovarianz der gepaarten Stichproben ist. Wie im Falle unabhängiger Gruppen führt diese Statistik auch bei Gültigkeit des Modells zu einer Überschätzung des Parameters d_D . Der erwartungstreue Schätzer ist hier: $\hat{d}_{D,s} = c(n-1) \cdot d_{D,s}$ (Morris & DeShon, 2002). Die Teststatistik des t -Tests für abhängige Stichproben lautet schließlich:

$$t = \frac{\bar{X}_D}{\sqrt{\frac{S_D^2}{n}}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2 + S_2^2 - 2S_{12}}{n}}},$$

wobei diese unter der Nullhypothese einer zentralen t -Verteilung mit $n - 1$ Freiheitsgraden und unter der Alternativhypothese einer nichtzentralen t -Verteilung mit $n - 1$ Freiheitsgraden und Nichtzentralitätsparameter $\delta = d_D \cdot \sqrt{n}$ folgt.

Voraussetzung dieses Tests ist die Normalverteilung der Differenzen bzw. des Mittelwertes der Differenzen. Die Homogenität der Varianzen spielt für die Validität des Verfahrens keine Rolle, wobei die Stärke dieses Tests wie im unabhängigen Fall bei Homogenität – sofern die Korrelation und die Summe der Varianzen konstant gehalten werden – am größten ist. Die Teststärke kann also nur unter Beachtung der Kovarianz korrekt bestimmt werden. Cohen aber empfiehlt als zu bevorzu-

gende Größe zur Beschreibung des Effekts den Parameter d aus der Analyse unabhängiger Gruppen (Cohen, 1988, S. 49), welcher bei Varianzhomogenität wie folgt mit d_D in Beziehung steht: $d = d_D \sqrt{2(1-\rho)}$. Bei Varianzhomogenität gilt demnach, daß $d = d_D$, wenn $\rho = .5$. Ferner nimmt d_D ein Minimum für $\rho = -1$ an und geht gegen unendlich, wenn $\rho \rightarrow 1$. Die Größe von d_D und somit die Teststärke wird wesentlich von der Korrelation der Variablen gesteuert. Daher ist d_D aber nur bedingt zur Beschreibung des (standardisierten) Lageunterschiedes geeignet. Zur flächenorientierten Interpretation von d in Einheiten der U -Statistiken oder zur Umrechnung in die punktbiserielle Korrelation sei auf Kapitel 2.1.1.1 verwiesen. Schließlich ist zu beachten, daß die genannte Beziehung in aller Regel nicht auf Ebene der Schätzer nutzbar gemacht werden kann, da hierfür die Identität der Varianzschätzungen erforderlich ist.

2.2.1.2 Cohens d' bei Abhängigkeit

Mit dem Ziel einer besseren Vergleichbarkeit mit Ergebnissen aus Untersuchungen unabhängiger Stichproben bietet sich neben der Varianz der Meßwertdifferenzen auch im abhängigen Fall die Varianz innerhalb der Gruppen zur Standardisierung an. Anders als von Cohen vorgeschlagen wird die Effektgröße d hier jedoch nicht weiter betrachtet, denn zum einen sollen auch in den Simulationen zum abhängigen Fall geeignete Statistiken für das Lagen-Skalen-Modell gefunden werden, für welches d jedoch nicht definiert ist. Zum anderen aber unterscheiden sich die Schätzwerte von d_s und d'_s bei Abhängigkeit nicht, da die gebündelte Varianz gleich dem Mittelwert der Varianzen ist, wenn $n_1 = n_2$. Somit stellt d_s bei Heterogenität ohnehin einen Schätzer für d' dar. Wie im unabhängigen Fall lautet der Parameter d' :

$$d' = \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}}.$$

Für die Beziehung zwischen d' und d_D gilt also folgendes:

$$d' = \sqrt{\frac{2(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12})}{\sigma_1^2 + \sigma_2^2}} d_D.$$

Die Beziehung zwischen d_D und d' hängt mit den Varianzen und der Kovarianz bzw. Korrelation wie folgt zusammen: Für den Fall $\sigma_1^2 = \sigma_2^2$ gilt $\sigma_D^2 = 2\sigma^2(1-\rho)$ und also ist $d_D = d'$, wenn $\rho = 0.5$ (s.o.). Sofern $d' \neq 0$ ist $|d_D| < |d'|$, wenn $\rho < .5$, und $d_D = d'$, falls $\rho = (\sigma_1^2 + \sigma_2^2)/(4\sigma_1\sigma_2)$. Wenn $\rho > (\sigma_1^2 + \sigma_2^2)/(4\sigma_1\sigma_2)$, ist $|d_D| > |d'|$. Die für die Gleichheit der Statistiken erforderliche Korrelation steigt, wenn das Verhältnis der Varianzen zunimmt. Wenn das Verhältnis der Varianzen den Wert 3.732² übersteigt, ist d_D grundsätzlich kleiner als d . Abbildung 22 gibt für einige Werte von ρ die Beziehung zwischen d' und d_D bei homogenen und heterogenen Varianzen ($\sigma_1^2/\sigma_2^2 = 9/1$) wieder. Hier wird deutlich, daß der Einfluß der Korrelation auf den Betrag von d_D abnimmt, wenn die Heterogenität der Varianzen größer wird.

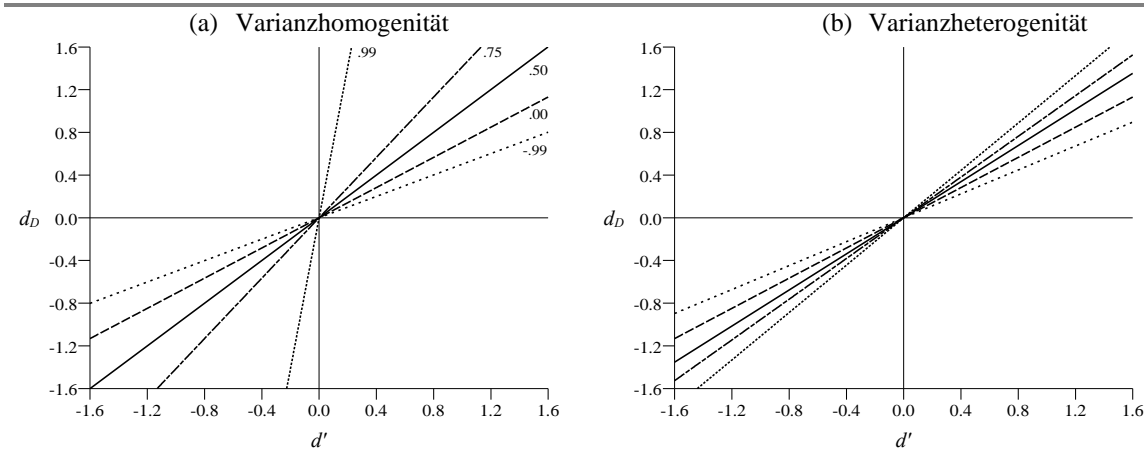


Abbildung 22. Beziehung zwischen d' und d_D in Abhängigkeit von der Korrelation bei homogenen und heterogenen Varianzen ($\sigma_1^2/\sigma_2^2 = 9/1$)

Als unkorrigierter Schätzer für d' kommt auch hier folgende Statistik in Frage:

$$d'_s = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2 + S_2^2}{2}}}.$$

Die Verwendung beider Gruppenvarianzen verspricht im Vergleich zur Standardisierung an nur einer Gruppenstreuung zwar auch im abhängigen Fall eine höhere Schätzgenauigkeit, jedoch sind die Verteilungseigenschaften von d'_s bisher nicht bekannt (Morris & DeShon, 2002; Viechtbauer, 2007). Insbesondere ist unbekannt, wie der Zugewinn an Freiheitsgraden mit der Korrelation verknüpft ist und wie der Nichtzentralitätsparameter lautet. Eine eher konservative Lösung des Problems besteht darin, die Beziehung zwischen d_D und d' , wie sie von Bird (2002) sowie Algina und Keselman (2003) für die Konstruktion asymptotischer Intervalle für d' genutzt wird, auch auf die Punktschätzung von d' anzuwenden:

$$\hat{d}'_{c1} = \sqrt{\frac{2(S_1^2 + S_2^2 - 2S_{12})}{S_1^2 + S_2^2}} \hat{d}_D,$$

wobei dieser Schätzer äquivalent ist zu $\hat{d}'_{c1} = c(n-1) \cdot d'_s$. Bei diesem Vorschlag wird also der Zugewinn an Freiheitsgraden durch die Berücksichtigung beider Gruppenvarianzen ignoriert. Bei kleinen Stichproben und $|\rho| < 1$ ist jedoch zu erwarten, daß diese Korrektur zu stark ist und zu einer Unterschätzung des Parameters führt.

Ausgehend von dem Umstand, daß die Freiheitsgrade von d'_s zwischen $n - 1$, wenn $|\rho| = 1$, und $2n - 2$, wenn $\rho = 0$, liegen, wird hier aufgrund von empirischen Voruntersuchungen folgende Erweiterung der Freiheitsgradenkorrektur nach Welch vorgeschlagen:

$$\hat{v} = \frac{n-1}{[q^2 + (1-q)^2]^{(1-r^2)}},$$

wobei $q = S_1^2 / (S_1^2 + S_2^2)$ und r die Korrelation sind. Als korrigierter Schätzer dient somit $\hat{d}'_{c_2} = c(\hat{v}) \cdot d'_s$. Tabelle 25 führt beispielhaft Ergebnisse der Voruntersuchung zum Bias der vorgestellten Statistiken zur Schätzung von d' auf.

Tabelle 25. Simulierte Erwartungswerte ($N_{MC} = 20,000$) unter varianzhomogenen und -heterogenen ($\sigma_1^2 / \sigma_2^2 = 9/1$) Normalverteilungen

n	ρ	Homogenität						Heterogenität					
		$d' = 0.8$			$d' = 1.6$			$d' = 0.8$			$d' = 1.6$		
		d'_s	\hat{d}'_{c_1}	\hat{d}'_{c_2}	d'_s	\hat{d}'_{c_1}	\hat{d}'_{c_2}	d'_s	\hat{d}'_{c_1}	\hat{d}'_{c_2}	d'_s	\hat{d}'_{c_1}	\hat{d}'_{c_2}
10	.0	.842	.769	.800	1.673	1.530	1.589	.854	.780	.795	1.706	1.560	1.588
	.4	.834	.763	.790	1.677	1.533	1.587	.856	.783	.795	1.724	1.576	1.601
	.8	.860	.786	.803	1.716	1.569	1.601	.867	.793	.799	1.745	1.595	1.604
20	.0	.809	.777	.791	1.632	1.566	1.596	.828	.795	.801	1.649	1.583	1.595
	.4	.821	.788	.802	1.636	1.571	1.598	.830	.797	.802	1.655	1.588	1.599
	.8	.827	.794	.801	1.654	1.588	1.602	.831	.798	.801	1.662	1.595	1.600

Vor allem bei $n = 10$ und fehlender oder geringer Korrelation liegt der (simulierte) Erwartungswert von \hat{d}'_{c_1} deutlich unterhalb des Parameters. Der Erwartungswert von \hat{d}'_{c_2} hingegen liegt bei allen hier gezeigten Bedingungen näher am Parameter d' .

2.2.1.3 Glass' Δ

Sowohl der Parameter d^{S_k} als auch der Schätzer d'_s der Effektgröße nach Glass sind im abhängigen Fall identisch mit denen bei unabhängigen Gruppen. Der Vorteil dieser Statistik im Vergleich zur Verwendung von Cohens d' liegt darin, daß die Verteilungseigenschaften auch bei abhängigen Stichproben bekannt sind (Becker, 1988; Morris & DeShon, 2002). Die Biaskorrektur lautet ebenfalls $c(n - 1)$.

2.2.1.4 Effektgrößen mit getrimmten Mittelwerten

Wie im unabhängigen Fall werden die getrimmten Effektgrößen analog zu den nicht getrimmten Statistiken definiert, wobei die jeweiligen Statistiken durch die getrimmten bzw. winsorisierten Varianten ersetzt werden (siehe Tabelle 26, S. 75). Während das getrimmte Mittel und die winsorisierte Varianz oben schon eingeführt wurden (vgl. Kap. 2.1.1.6), ist hier zudem die winsorisierte Kovarianz zu betrachten. Bei abhängigen Gruppen kann die Winsorisierung wie folgt notiert werden:

$$X_{(w)ik} = \begin{cases} (1-q)X_{(g+1)k} + qX_{(g+2)k} & \text{wenn } X_{ik} \leq X_{(g+1)k} \\ X_{ik} & \text{wenn } X_{(g+1)k} < X_{ik} < X_{(n-g)k} \\ (1-q)X_{(n-g)k} + qX_{(n-g-1)k} & \text{wenn } X_{ik} \geq X_{(n-g)k} \end{cases}$$

wobei der Trimmanteil allgemein $0 < \gamma < .5$, hier jedoch stets $.2$, für jede Seite der geordneten Statistik $X_{(1)k} \leq X_{(2)k} \leq \dots \leq X_{(n)k}$ von Gruppe k , $g = [\gamma n]$ und $q = \gamma n - g$ sind. Der Schätzer der winsorisierten Kovarianz entspricht der herkömmlichen Statistik der Kovarianz berechnet über die winsorisierten Werte $X_{(w)ik}$:

$$S_{w12} = \frac{1}{n} \sum (X_{(w)i1} - \bar{X}_{(w)1})(X_{(w)i2} - \bar{X}_{(w)2}).$$

Die winsorisierte Korrelation lautet schließlich $r_w = S_{w12} / (S_{w1} S_{w2})$.

2.2.2 Intervallschätzer

Für d_D kann unter der Normalverteilung ein exaktes nichtzentrales Konfidenzintervall berechnet werden. Für die anderen Effektgrößen lassen sich bisher hingegen nur approximative Intervalle konstruieren. Wie bei unverbundenen Stichproben wird auf die Untersuchung der zentralen Intervalle verzichtet (vgl. Viechtbauer, 2007; Wu & Jiang, 2007), da sie gegenüber dem nichtzentralen Vorgehen keine Vorteile hinsichtlich der Robustheit bieten.

2.2.2.1 Exakter Schätzer für d_D

Die Konstruktion des exakten, nichtzentralen Intervalls folgt der in Kapitel 2.1.2.2 näher beschriebenen Methode, bei welcher im ersten Schritt anhand der beobachteten Teststatistik ein Intervall für den Nichtzentralitätsparameter der t -Verteilung bestimmt und anschließend in die Einheit der Effektgröße umgerechnet wird. Beim t -Test für abhängige Stichproben lautet die Beziehung $\delta = d_D \cdot \sqrt{n}$. Die Intervallgrenzen für den Nichtzentralitätsparameter werden also durch \sqrt{n} dividiert, um das Intervall für d_D zu erhalten.

2.2.2.2 Nichtzentrale Schätzer für d^{S_k} und d'

Für die Effektgröße d^{S_k} nach Glass kann bei Annahme homogener Varianzen folgende Beziehung zum Nichtzentralitätsparameter einer t -Verteilung mit $n - 1$ Freiheitsgraden angegeben werden: $\delta = d^{S_k} \cdot \sqrt{n/2(1-\rho)}$ (Becker, 1988; Morris & DeShon, 2002). Grundlage dieser Formel ist folgende Umrechnung der Effektgrößen: $d_D = d^{S_k} / \sqrt{2(1-\rho)}$. Da die entsprechend konstruierten Intervalle jedoch im Falle heterogener Varianzen das nominale Niveau nicht einhalten können, denn dann ist $d_D \neq d^{S_k} / \sqrt{2(1-\rho)}$, wird hier nach einem Vorschlag von Algina, Keselman und Penfield (2005b) folgende Beziehung zugrundegelegt: $d_D = d^{S_k} \sqrt{\sigma_k^2 / (\sigma_1^2 + \sigma_2^2 - 2\sigma_{12})}$. Konkret werden also die exakten Intervallgrenzen für d_D anhand dieser Beziehung und der entsprechenden Stichprobenkennwerte in ein Intervall für d^{S_k} umgerechnet. Aufgrund der zusätzlichen Variabilität, die durch die erforderliche Schätzung der Varianzen und der Kovarianz entsteht, sind die nichtzentralen Intervalle für d^{S_k} auch unter der Normalverteilung keine exakten, sondern approximative Intervalle (Viechtbauer, 2007).

Bei der Schätzung eines Intervalls für d' besteht zudem das Problem, daß die Verteilungseigenschaften von d'_s bei abhängigen Stichproben nicht bekannt sind (Morris & DeShon, 2002, S. 111). Bird (2002) sowie Algina und Keselman (2003) haben Intervallschätzer vorgestellt, bei welchen die Beziehung $d' = d_D \sqrt{2(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}) / (\sigma_1^2 + \sigma_2^2)}$ zwischen den Effektgrößen auf die Intervallgrenzen eines Schätzers für d_D angewendet wird. Da die Überdeckungsrate des von Bird vorgeschlagenen Intervalls auf Basis der zentralen t -Verteilung mit zunehmendem Effekt jedoch selbst unter der Normalverteilung drastisch abnimmt (Algina & Keselman, 2003, S. 541), wird hier nur der nichtzentrale Ansatz von Algina und Keselman verfolgt. Wie bei der Schätzung von d^{S_k} wird also die Intervallschätzung für d_D mit den entsprechenden Stichprobenkennwerten in ein Intervall für d' umgerechnet (in der Auswertung als NC_1 gekennzeichnet). Da aber die Freiheitsgrade von d'_s nicht gleich $n - 1$ sind, sondern in Abhängigkeit von der Korrelation zwischen $n - 1$ und $2n - 2$ liegen,

wird hier nach dem ansonsten gleichen Prinzip zudem ein Intervall vorgeschlagen, bei welchem die in Kapitel 2.2.1.2 beschriebene Freiheitsgradenkorrektur eingesetzt wird (Abkürzung: NC_2).

2.2.3 Fragestellung und ausgewählte Statistiken

Auch im Fall abhängiger Stichproben folgt die Auswahl der Statistiken dem Ziel, diese hinsichtlich ihrer Eignung für das allgemeine Lagen-Skalen-Modell einzuschätzen. Somit fällt aber bereits eine Vielzahl von Intervallschätzern aus der Betrachtung heraus, da diese explizit homogene Varianzen voraussetzen. So sind sämtliche, insgesamt 21 zentralen Schätzer, die Viechtbauer (2007) verglichen hat, ebenso wie die von Wu und Jiang (2007) untersuchten Verfahren, für heterogene Varianzen nicht geeignet.

Die einzige Studie, in welcher für verbundene Stichproben auch der Fall heterogener Varianzen berichtet wird, stammt von Algina, Keselman und Penfield (2005b). In dieser Arbeit werden die vorgestellten nichtzentralen Schätzer für d' (hier NC_1 genannt) und d^{S_k} mit KQ- und robustifizierten Statistiken (20%-getrimmt/winsorisiert) simuliert und jeweils dem Verhalten des Perzentil-Bootstraps gegenübergestellt. Schätzer für d_D werden in dieser Arbeit nur insofern behandelt, als sie die Grundlage für die Konstruktion der genannten nichtzentralen Intervalle bilden. Obwohl das von den Autoren vorgeschlagene Verfahren zur Intervallschätzung von d' auch den Fall heterogener Varianzen einschließt (S. 242), da nicht allein die Korrelation, sondern die Varianzen und die Kovarianz getrennt in die Umrechnung des nichtzentralen Intervalls eingehen, empfehlen die Autoren bei Varianzheterogenität die Verwendung einer Effektgröße vom Glass'schen Typus. Für die Schätzung von d' werden daher nur Simulationen unter homogenen Verteilungen berichtet. Als Verteilungsmodelle werden neben der Normalverteilung drei $g-h$ -Verteilungen realisiert, welche auch in der vorliegenden Arbeit verwendet werden, nämlich $g-h-1$, $g-h-2$ und $g-h-3$. Die Ergebnisse der Studie von Algina, Keselman und Penfield (2005b) lassen sich wie folgt zusammenfassen: Sowohl die nichtzentralen als auch die Perzentil-Verfahren mit KQ-Schätzern halten nur unter der Normalverteilung das nominale Niveau ein. Unter den $g-h$ -Verteilungen liegt die Überdeckungswahrscheinlichkeit dieser Verfahren insbesondere bei größeren Parameterwerten deutlich unterhalb der nominalen Wahrscheinlichkeit. In Verbindung mit robustifizierten Schätzern halten diese Verfahren das nominale Niveau hingegen sehr viel besser ein. Da aber unter schiefen Verteilungen mit wachsendem Parameterwert und zunehmender Korrelation das nichtzentrale Intervall auch mit robustifizierten Schätzern zur Unterschreitung der nominalen Wahrscheinlichkeit neigt, empfehlen die Autoren für den homogenen und besonders für den heterogenen Fall das Perzentil-Bootstrap.

In der vorliegenden Arbeit wird nun die Simulationsstudie von Algina et al. in folgenden Punkten erweitert: (1) Neben d' und d^{S_k} wird als weitere Effektgröße d_D berücksichtigt. (2) Es werden nicht nur Intervallschätzer, sondern auch Punktschätzer untersucht. (3) Als weiteres Kriterium für die Güte der Intervallschätzer wird neben der Überdeckungswahrscheinlichkeit auch die Intervalllänge herangezogen. (4) Alle Schätzer werden auch unter heterogenen Verteilungen simuliert. (5) Zusätzlich zu den Verteilungen mit Tailstärken oberhalb der Normalverteilung werden auch Verteilungen mit kurzen Tails modelliert. (6) Neben stetigen Verteilungen wird auch ein diskretes Modell simuliert. (7) Es werden neben dem Perzentil-Bootstrap auch das BCA- und das Perzentil- t -Verfahren zur Intervallschätzung eingesetzt. Die wesentliche Fragestellung der folgenden Studie lautet ganz analog zum unabhängigen Fall, ob mit nichtzentralen und Bootstrap-Verfahren auch im varianzheterogenen Fall robuste Intervalle für die Schätzung von d_D , d' und d^{S_k} sowie deren robustifizierten Varianten konstruiert werden können.

2.2.3.1 Punktschätzer

Tabelle 26 faßt alle für die folgenden Simulationen ausgewählten Parameter und Punktschätzer der Effektgrößen bei zwei verbundenen Stichproben zusammen.

Tabelle 26. Parameter und Punktschätzer der Effektgrößen bei zwei abhängigen Gruppen¹

KQ		(20%-)robustifiziert	
Parameter	Schätzer	Parameter	Schätzer
$d_D = \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}}$	$d_{D,s} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_1^2 + S_2^2 - 2S_{12}}}$	$d_{D,R} = \frac{\mu_{t1} - \mu_{t2}}{\sqrt{\sigma_{w1}^2 + \sigma_{w2}^2 - 2\sigma_{w12}}}$	$d_{D,s,R} = \frac{\bar{X}_{t1} - \bar{X}_{t2}}{\sqrt{S_{w1}^2 + S_{w2}^2 - 2S_{w12}}}$
	$\hat{d}_D = c(n-1)d_{D,s}$		$\hat{d}_{D,R} = c(h-1)d_{D,s,R}$
	$d_{D,s}^{Boot} = 2d_{D,s} - \frac{1}{B} \sum d_{D,s,b}^*$		$d_{D,s,R}^{Boot} = 2d_{D,s,R} - \frac{1}{B} \sum d_{D,s,R,b}^*$
	$\hat{d}_D^{Boot} = 2\hat{d}_D - \frac{1}{B} \sum \hat{d}_{D,b}^*$		$\hat{d}_{D,R}^{Boot} = 2\hat{d}_{D,R} - \frac{1}{B} \sum \hat{d}_{D,R,b}^*$
$d'_s = \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}}$	$d'_s = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2 + S_2^2}{2}}}$	$d'_{R} = \frac{\mu_{t1} - \mu_{t2}}{\sqrt{\frac{\sigma_{w1}^2 + \sigma_{w2}^2}{2}}}$	$d'_{s,R} = \frac{\bar{X}_{t1} - \bar{X}_{t2}}{\sqrt{\frac{S_{w1}^2 + S_{w2}^2}{2}}}$
	$\hat{d}'_{cl} = c(n-1)\hat{d}_D$		$\hat{d}'_{cl,R} = c(h-1)\hat{d}_{D,R}$
	$\hat{d}'_{c2} = c \left(\frac{n-1}{[q^2 + (1-q)^2]^{(1-r^2)}} \right) d'_s$ mit $q = S_1^2 / (S_1^2 + S_2^2)$		$\hat{d}'_{c2,R} = c \left(\frac{h-1}{[q^2 + (1-q)^2]^{(1-r_w^2)}} \right) d'_{s,R}$ mit $q = S_{w1}^2 / (S_{w1}^2 + S_{w2}^2)$
	$d_s'^{Boot} = 2d'_s - \frac{1}{B} \sum d_{s,b}'^*$		$d_{s,R}'^{Boot} = 2d'_{s,R} - \frac{1}{B} \sum d_{s,R,b}'^*$
	$\hat{d}'_{cl}^{Boot} = 2\hat{d}'_{cl} - \frac{1}{B} \sum \hat{d}'_{cl,b}^*$		$\hat{d}'_{cl,R}^{Boot} = 2\hat{d}'_{cl,R} - \frac{1}{B} \sum \hat{d}'_{cl,R,b}^*$
	$\hat{d}'_{c2}^{Boot} = 2\hat{d}'_{c2} - \frac{1}{B} \sum \hat{d}'_{c2,b}^*$		$\hat{d}'_{c2,R}^{Boot} = 2\hat{d}'_{c2,R} - \frac{1}{B} \sum \hat{d}'_{c2,R,b}^*$
$d^{S_k} = \frac{\mu_1 - \mu_2}{\sigma_k}$	$d_s^{S_k} = \frac{\bar{X}_1 - \bar{X}_2}{S_k}$	$d_R^{S_k} = \frac{\mu_{t1} - \mu_{t2}}{\sigma_{wk}}$	$d_{s,R}^{S_k} = \frac{\bar{X}_{t1} - \bar{X}_{t2}}{S_{wk}}$
	$\hat{d}^{S_k} = c(n-1)d_s^{S_k}$		$\hat{d}_R^{S_k} = c(h-1)d_{s,R}^{S_k}$
	$d_s^{S_k,Boot} = 2d_s^{S_k} - \frac{1}{B} \sum d_{s,b}^{S_k,*}$		$d_{s,R}^{S_k,Boot} = 2d_{s,R}^{S_k} - \frac{1}{B} \sum d_{s,R,b}^{S_k,*}$
	$\hat{d}^{S_k,Boot} = 2\hat{d}^{S_k} - \frac{1}{B} \sum \hat{d}_b^{S_k,*}$		$\hat{d}_R^{S_k,Boot} = 2\hat{d}_R^{S_k} - \frac{1}{B} \sum \hat{d}_{R,b}^{S_k,*}$

¹ Die t -basierte Biaskorrektur lautet: $c(m) = \Gamma(m/2) / [(m/2)^{-0.5} \Gamma((m-1)/2)]$.

2.2.3.2 Intervallschätzer

Die Notation der jeweils sieben Intervallschätzer für die Parameter d_D , $d_{D,R}$, d^{S_k} und $d_R^{S_k}$ folgt dem bereits in Kapitel 2.1.3.2 dargestellten Schema. Da für die Schätzer der Parameter d' und d'_R jeweils zwei Formeln zur Berechnung der Freiheitsgrade untersucht werden, ergeben sich insgesamt je zwölf Varianten, die am Beispiel von d' in Tabelle 27 beschrieben werden.

Tabelle 27. Notation der ausgewählten Intervallschätzer am Beispiel von d'

Intervallschätzer	Beschreibung des Verfahrens
Perc (d'_s)	Perzentil-Bootstrap mit nichtkorrigiertem Punktschätzer
BCA (d'_s)	BCA-Bootstrap mit nichtkorrigiertem Punktschätzer
Perc- t_1 (d'_s)	Perzentil- t -Bootstrap mit nichtkorrigiertem Punktschätzer und $S(d'_s)_1$
Perc- t_2 (d'_s)	Perzentil- t -Bootstrap mit nichtkorrigiertem Punktschätzer und $S(d'_s)_2$
Perc (\hat{d}'_{c1})	Perzentil-Bootstrap mit korrigiertem Punktschätzer \hat{d}'_{c1}
BCA (\hat{d}'_{c1})	BCA-Bootstrap mit korrigiertem Punktschätzer \hat{d}'_{c1}
Perc- t (\hat{d}'_{c1})	Perzentil- t -Bootstrap mit korrigiertem Punktschätzer \hat{d}'_{c1} und $S(\hat{d}'_1)$
NC ₁ (d'_s)	Nichtzentrales Intervall mit $\nu = n - 1$
Perc (\hat{d}'_{c2})	Perzentil-Bootstrap mit korrigiertem Punktschätzer \hat{d}'_{c2}
BCA (\hat{d}'_{c2})	BCA-Bootstrap mit korrigiertem Punktschätzer \hat{d}'_{c2}
Perc- t (\hat{d}'_{c2})	Perzentil- t -Bootstrap mit korrigiertem Punktschätzer \hat{d}'_{c2} und $S(\hat{d}'_2)$
NC ₂ (d'_s)	Nichtzentrales Intervall mit $\hat{\nu} = (n-1)/[q^2 + (1-q)^2]^{(1-r^2)}$

Die für das Perzentil- t -Bootstrap erforderlichen Standardfehlerschätzer sind wiederum beispielhaft für die KQ-Statistiken in Tabelle 28 aufgeführt. Die robustifizierten Schätzer ergeben sich durch Einsetzen der entsprechenden getrimmten und winsorisierten Statistiken.

Tabelle 28. Standardfehlerschätzer der Effektgrößen bei verbundenen Stichproben

$$S(d_{D,s}) = \sqrt{\frac{1}{n} \cdot \frac{n-1}{n-3} \cdot (1 + n \cdot d_{D,s}^2) - \frac{d_{D,s}^2}{c(n-1)}}$$

$$S(\hat{d}_D) = c(n-1)S(d_{D,s})$$

$$S(d'_s)_1 = \sqrt{\frac{2(S_1^2 + S_2^2 - 2S_{12})}{S_1^2 + S_2^2}} \cdot S(d_{D,s})$$

$$S(\hat{d}'_1) = c(n-1)S(d'_s)_1$$

$$S(d'_s)_2 = \sqrt{\frac{2(S_1^2 + S_2^2 - 2S_{12})}{S_1^2 + S_2^2}} \sqrt{n} \cdot S_{t(\hat{\nu}, \hat{\delta})} \text{ mit } \hat{\nu} = \frac{n-1}{[q^2 + (1-q)^2]^{(1-r^2)}}, q = S_1^2 / (S_1^2 + S_2^2) \text{ und } \hat{\delta} = \frac{d_{D,s}}{\sqrt{n}}$$

$$S(\hat{d}'_2) = c(\hat{\nu})S(d'_s)_2 \text{ mit } \hat{\nu} \text{ wie oben}$$

$$S(d_s^{S_k}) = \sqrt{\frac{S_1^2 + S_2^2 - 2S_{12}}{S_k^2}} \cdot S(d_{D,s})$$

$$S(\hat{d}^{S_k}) = c(n-1)S(d_s^{S_k})$$

2.2.4 Simulationsmethode

Hinsichtlich der verwendeten Verteilungen, der Robustheitskriterien und der Anzahl der Simulationsdurchgänge folgt die Studie zu verbundenen Stichproben der Gestaltung von Studie 1 (vgl. Kapitel 2.1.4).

2.2.4.1 Erzeugung der korrelierten Pseudozufallszahlen

Über folgende Transformation, die ein Spezialfall der Cholesky-Faktorisierung für zwei Variablen ist (Moonan, 1957), lassen sich aus unabhängigen, standardnormalverteilten Zufallsvariablen X_1 und X_2 bivariat standardnormalverteilte Variablen Z_1 und Z_2 mit Korrelation ρ erzeugen: $Z_1 = X_1$ und $Z_2 = \rho \cdot X_1 + \sqrt{1-\rho^2} \cdot X_2$. Die so gewonnene bivariate Standardnormalverteilung ist jeweils Ausgangspunkt für alle hier simulierten bivariaten Verteilungen. Zur Erzeugung bivariater g - h -Verteilungen werden die korrelierten Randverteilungen Z_1 und Z_2 den in Kapitel 2.1.4.2 beschriebenen Transformationen unterzogen; zur Erzeugung korrelierter $(0, 1)$ -Gleichverteilungen werden Z_1 und Z_2 nach der Inversionsmethode in die Verteilungsfunktion der Standardnormalverteilung eingesetzt. Die so erzeugten Gleichverteilungen sind wiederum Grundlage für die negativen hypergeometrischen Verteilungen.

Im Falle der Gleichverteilung ist die Beziehung zwischen den linearen Zusammenhängen bekannt. Um Gleichverteilungen mit ρ_G zu erzeugen, muß für die Normalverteilungen ρ_N gewählt werden: $\rho_N = 2 \sin[(\rho_G \pi)/6]$. Für die g - h - und die negativen hypergeometrischen Verteilungen sind die Beziehungen formal nicht bekannt und werden in der Literatur zumeist empirisch bestimmt (z.B. Algina, Keselman & Penfield, 2005b). Die erforderlichen ρ_N -Werte sind hier über sukzessive Approximationen anhand von jeweils 20,000,000 Fällen bis auf die dritte Dezimalstelle genau berechnet worden (vgl. Anhang B). Um beispielsweise eine bivariate g - h - $(0; 0.225)$ -Verteilung mit $\rho = .800$ zu erzeugen, muß die zugrundeliegende Normalverteilung eine Korrelation von $\rho_N = .8337$ (vgl. Tabelle B-2) aufweisen.

2.2.4.2 Datenmodelle

Den Randverteilungen der folgenden Simulationen liegen dieselben Datenmodelle wie im unabhängigen Fall zugrunde, wobei aber X_1 und X_2 nun bivariat verteilt sind mit Korrelationsparameter $\rho(X_1, X_2)$. Für die stetigen (Rand-)Verteilungen gilt somit:

$$X_{1i} \sim F\left(\frac{x - \mu_1}{\sigma_1}\right) \text{ und } X_{2i} \sim F\left(\frac{x - \mu_2}{\sigma_2}\right).$$

Für die negativen hypergeometrischen Verteilungen gilt:

$$X_{1i} \sim H(n = 10; a_1; b_1) \text{ und } X_{2i} \sim H(n = 10; a_2; b_2).$$

2.2.4.3 Untersuchungsdesign

Das Untersuchungsschema von Studie 1 wurde hier weitgehend beibehalten, die unabhängigen Variablen Verteilung, N , Effektgröße und Verhältnis der Varianzen wurden mit denselben Stufen realisiert. Als neue Variable kommt die Korrelation zwischen den Verteilungen mit drei Stufen $\{0, .4, .8\}$ hinzu. Diese Variablen wurden wie folgt miteinander gekreuzt:

$$\begin{array}{c}
 \text{Verteilung} \\
 \left(\begin{array}{c}
 \text{Normal} \\
 \text{Gleich} \\
 g-h-1 \\
 g-h-2 \\
 g-h-3 \\
 g-h-4 \\
 \text{Hyper}
 \end{array} \right)
 \end{array}
 \times
 \begin{array}{c}
 N \\
 \left(\begin{array}{c}
 40 \\
 120
 \end{array} \right)
 \end{array}
 \times
 \begin{array}{c}
 \{\sigma_1; \sigma_2; \rho\} \\
 \left(\begin{array}{c}
 1:1; .0 \\
 1:1; .4 \\
 1:1; .8 \\
 3:1; .0 \\
 3:1; .4 \\
 3:1; .8
 \end{array} \right)
 \end{array}
 \times
 \begin{array}{c}
 d' \\
 \left(\begin{array}{c}
 0 \\
 0.4 \\
 0.8 \\
 1.2 \\
 1.6
 \end{array} \right)
 \end{array}
 .$$

Insgesamt umfaßt die Studie 2 zur Untersuchung der standardisierten Lagedifferenzen bei verbundenen Stichproben also 420 Bedingungen.

2.2.5 Ergebnisse der Studie 2

Der folgende Ergebnisteil zu Studie 2 ist analog zu den dargestellten Ergebnissen von Studie 1 aufgebaut, d.h. es werden jeweils getrennt nach den verschiedenen Parametern zuerst die Punktschätzer und anschließend die Intervallschätzer im Hinblick auf die in Kapitel 2.1.5.1 und Kapitel 2.1.5.2 definierten Kriterien untersucht.

2.2.5.1 Punktschätzer

Wie dem Anhang D (vgl. Tabelle D-1 bis Tabelle D-16) zu entnehmen ist, hängen die Verzerrungen sowohl der KQ-Schätzer als auch der robustifizierten Statistiken weder von der Varianzheterogenität noch von der Korrelation der zugrundeliegenden Verteilungen in wesentlicher Weise ab. Zur Vereinfachung der Darstellung werden daher die Verzerrungen im folgenden nur nach Verteilung und Stichprobengröße getrennt abgebildet. Auf die direkte Gegenüberstellung von herkömmlicher und robustifizierter Punktschätzung wird hier verzichtet; ein Vergleich kann anhand der Tabellen 29 bis 34 vorgenommen werden (siehe auch Kapitel 2.1.5.1.5).

2.2.5.1.1 Schätzer für d_D

Von den vier ausgewählten Schätzern weist insgesamt der auf Basis des Bootstraps korrigierte Schätzer \hat{d}_D^{Boot} im Mittel die kleinste Verzerrung sowie auch den kleinsten quadratischen Fehler auf (vgl. Tabelle 29).

Tabelle 29. $\overline{B_{MC}^2}$ und $\overline{MSE_{MC}}$ der Schätzer für d_D

	$n = 20$		$n = 60$		Gesamt	
	$\overline{B_{MC}^2}$	$\overline{MSE_{MC}}$	$\overline{B_{MC}^2}$	$\overline{MSE_{MC}}$	$\overline{B_{MC}^2}$	$\overline{MSE_{MC}}$
$d_{D,s}$	0.014	0.141	0.003	0.045	0.008	0.093
\hat{d}_D	0.007	0.125	0.002	0.043	0.005	0.084
$d_{D,s}^{Boot}$	0.002	0.123	0.001	0.045	<u>0.002</u>	0.084
\hat{d}_D^{Boot}	0.003	0.115	0.001	0.044	<u>0.002</u>	<u>0.079</u>

Anmerkung: Unterstreichungen kennzeichnen die besten Resultate je Kriterium

Dieser Vorteil geht vor allem auf das Verhalten bei tailstarken Verteilungen zurück. Bei diesen Verteilungen ($g-h-1$ und $g-h-3$) führen die anderen Schätzer zu deutlich größeren Überschätzungen des Parameters (vgl. Abbildung 23). Unter allen anderen Verteilungen führt \hat{d}_D^{Boot} mit zunehmenden

dem Parameterwert jedoch zu einer größeren Unterschätzung. Unter den Verteilungen mit normalen und kurzen Tails schneidet insgesamt der auf Basis der t -Verteilung korrigierte Schätzer \hat{d}_D am besten ab.

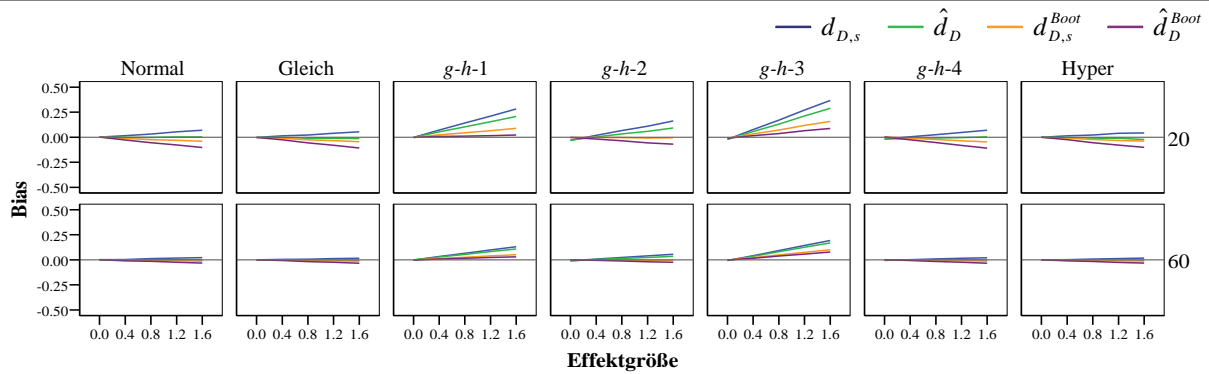


Abbildung 23. Mittlere Verzerrungen der Schätzer für d_D getrennt nach Stichprobengröße

2.2.5.1.2 Schätzer für d'

Auch für die Schätzung von d' lassen insgesamt die Verfahren mit Bootstrap-Biaskorrektur die geringsten Verzerrungen erkennen (vgl. Tabelle 30), wobei diese Vorteile wiederum auf ihr Verhalten unter den tailstarken $g-h$ -Verteilungen zurückgehen (siehe Abbildung 24).

Tabelle 30. $\overline{B_{MC}^2}$ und $\overline{MSE_{MC}}$ der Schätzer für d'

	$n = 20$		$n = 60$		Gesamt	
	$\overline{B_{MC}^2}$	$\overline{MSE_{MC}}$	$\overline{B_{MC}^2}$	$\overline{MSE_{MC}}$	$\overline{B_{MC}^2}$	$\overline{MSE_{MC}}$
d'_s	0.009	0.121	0.002	0.040	0.006	0.080
\hat{d}'_{c1}	0.005	0.108	0.001	0.038	0.003	0.073
\hat{d}'_{c2}	0.006	0.111	0.002	0.038	0.004	0.075
d'^{Boot}_s	0.001	0.110	0.000	0.040	<u>0.001</u>	0.075
\hat{d}'^{Boot}_{c1}	0.003	0.103	0.000	0.039	0.002	<u>0.071</u>
\hat{d}'^{Boot}_{c2}	0.002	0.105	0.000	0.039	<u>0.001</u>	0.072

Anmerkung: Unterstreichungen kennzeichnen die besten Resultate je Kriterium

Unter den Verteilungen mit schwächeren Tails weist \hat{d}'_{c1} die besten Schätzeigenschaften auf (vgl. Tabelle D-3 und Tabelle D-4), wobei auch hier die Bootstrap-Schätzer mit steigendem Parameter zu einer zunehmenden Unterschätzung führen.

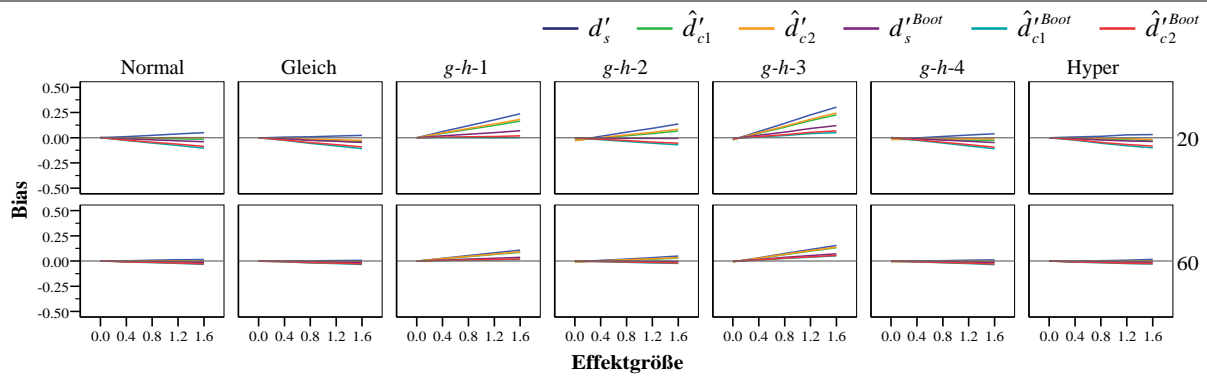


Abbildung 24. Mittlere Verzerrungen der Schätzer für d' getrennt nach Stichprobengröße

2.2.5.1.3 Schätzer für d^{S_1} und d^{S_2}

Das für d_D und d' beschriebene Verhalten der verschiedenen Schätzarten kann auch bei der Schätzung für d^{S_k} beobachtet werden: Die Bootstrap-Schätzer zeigen insgesamt eine geringere Verzerrung und führen im Gegensatz zu den klassischen Schätzern tendentiell zu einer Unterschätzung der Parameter (vgl. Tabelle 31 und Abbildung 25).

Tabelle 31. $\overline{B_{MC}^2}$ und $\overline{MSE_{MC}}$ der Schätzer für d^{S_1} und d^{S_2}

	$n = 20$		$n = 60$		Gesamt	
	$\overline{B_{MC}^2}$	$\overline{MSE_{MC}}$	$\overline{B_{MC}^2}$	$\overline{MSE_{MC}}$	$\overline{B_{MC}^2}$	$\overline{MSE_{MC}}$
$d_s^{S_1}$	0.013	0.134	0.003	0.042	0.008	0.088
$\hat{d}_s^{S_1}$	0.007	0.119	0.002	0.040	0.005	0.080
$d_s^{S_1,Boot}$	0.002	0.117	0.001	0.042	<u>0.001</u>	0.079
$\hat{d}_s^{S_1,Boot}$	0.003	0.109	0.001	0.040	0.002	<u>0.075</u>
$d_s^{S_2}$	0.019	0.199	0.004	0.059	0.011	0.129
$\hat{d}_s^{S_2}$	0.011	0.177	0.003	0.056	0.007	0.116
$d_s^{S_2,Boot}$	0.003	0.172	0.001	0.057	<u>0.002</u>	0.115
$\hat{d}_s^{S_2,Boot}$	0.003	0.159	0.001	0.056	<u>0.002</u>	<u>0.107</u>

Anmerkung: Unterstreichungen kennzeichnen die besten Resultate je Kriterium

Insgesamt läßt sich für die Schätzung von Effektgrößen mit kleinsten Quadratschätzern also festhalten, daß im Falle starker Tails eine Schätzung auf Basis des Bootstraps in Betracht zu ziehen ist, während bei normalen und kurzen Tails Schätzungen mit t -basierter Korrektur vorzuziehen sind. Die simulierten Erwartungswerte der Bootstrap-biaskorrigierten Schätzer liegen in aller Regel unterhalb der anhand der t -Verteilung korrigierten Schätzer und letztere liegen – da $c(m) < 1$ ist – unterhalb derjenigen der nichtkorrigierten Schätzer.

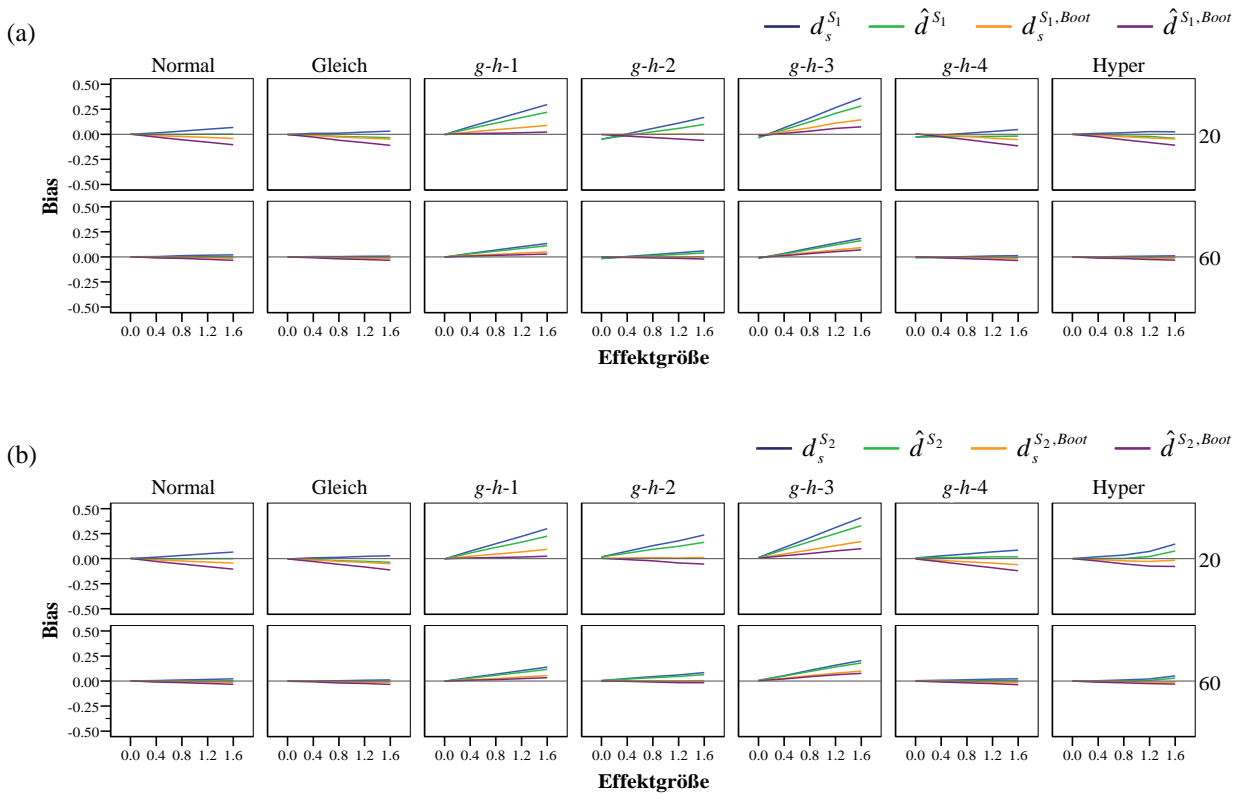


Abbildung 25. Mittlere Verzerrungen der Schätzer für d^{S_1} und d^{S_2} getrennt nach Stichprobengröße

2.2.5.1.4 Schätzer für $d_{D,R}$

Anders als bei den KQ-Schätzern schneidet bei der Schätzung von $d_{D,R}$ insgesamt der t -basierte Schätzer $\hat{d}_{D,R}$ am besten ab (Tabelle 32). Der Bootstrap-biaskorrigierte Schätzer $d_{D,s,R}^{Boot}$ weist zwar einen ebenso geringen $\overline{B_{MC}^2}$ auf, der $\overline{MSE_{MC}}$ ist jedoch deutlich größer.

Tabelle 32. $\overline{B_{MC}^2}$ und $\overline{MSE_{MC}}$ der Schätzer für $d_{D,R}$

	$n = 20$		$n = 60$		Gesamt	
	$\overline{B_{MC}^2}$	$\overline{MSE_{MC}}$	$\overline{B_{MC}^2}$	$\overline{MSE_{MC}}$	$\overline{B_{MC}^2}$	$\overline{MSE_{MC}}$
$d_{D,s,R}$	0.008	0.184	0.001	0.046	0.004	0.115
$\hat{d}_{D,R}$	0.001	0.154	0.001	0.044	<u>0.001</u>	<u>0.099</u>
$d_{D,s,R}^{Boot}$	0.002	0.201	0.001	0.051	<u>0.001</u>	0.126
$\hat{d}_{D,R}^{Boot}$	0.004	0.177	0.001	0.049	0.002	0.113

Anmerkung: Unterstreichungen kennzeichnen die besten Resultate je Kriterium

Wie Abbildung 26 zeigt, ergeben sich im Vergleich dieser beiden Schätzer keine spezifischen Vorteile in Abhängigkeit von der zugrundeliegenden Verteilung, so daß aufgrund der vorliegenden Simulationen die t -basierte Korrektur $\hat{d}_{D,R}$ empfohlen werden kann.

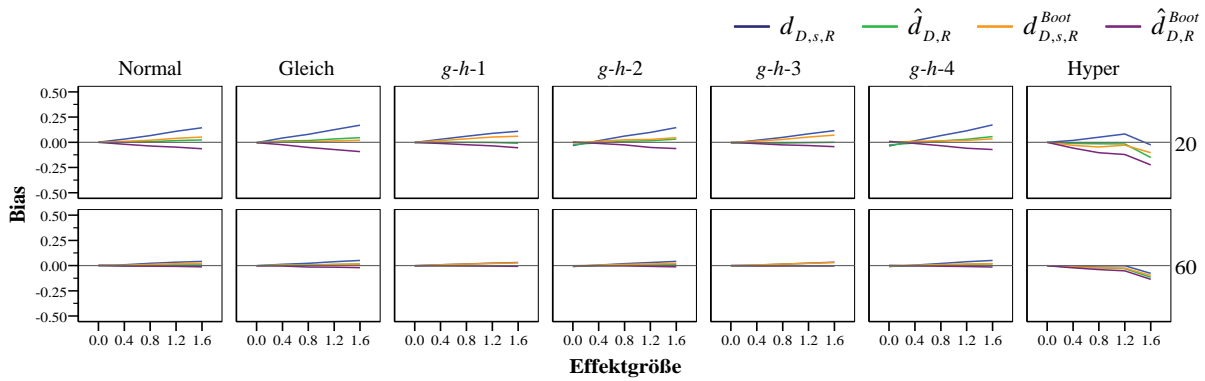


Abbildung 26. Mittlere Verzerrungen der Schätzer für $d_{D,R}$ getrennt nach Stichprobengröße

2.2.5.1.5 Schätzer für d'_R

Bis auf den nichtkorrigierten Schätzer $d'_{s,R}$ weisen hier alle Schätzer insgesamt einen gemittelten quadratischen Bias von jeweils .001 auf (vgl. Tabelle 33), die t -korrigierten Schätzer haben jedoch wiederum einen geringeren \overline{MSE}_{MC} als die Bootstrap-Korrekturen. Den insgesamt kleinsten quadratischen Fehler besitzt $\hat{d}'_{c1,R}$.

Tabelle 33. $\overline{B^2}_{MC}$ und \overline{MSE}_{MC} der Schätzer für d'_R

	$n = 20$		$n = 60$		Gesamt	
	$\overline{B^2}_{MC}$	\overline{MSE}_{MC}	$\overline{B^2}_{MC}$	\overline{MSE}_{MC}	$\overline{B^2}_{MC}$	\overline{MSE}_{MC}
$d'_{s,R}$	0.007	0.164	0.001	0.043	0.004	0.103
$\hat{d}'_{c1,R}$	0.001	0.137	0.000	0.041	<u>0.001</u>	<u>0.089</u>
$\hat{d}'_{c2,R}$	0.001	0.144	0.000	0.041	<u>0.001</u>	0.093
$d'^{Boot}_{s,R}$	0.002	0.177	0.001	0.046	<u>0.001</u>	0.112
$\hat{d}'^{Boot}_{c1,R}$	0.002	0.154	0.001	0.044	<u>0.001</u>	0.099
$\hat{d}'^{Boot}_{c2,R}$	0.001	0.161	0.001	0.045	<u>0.001</u>	0.103

Anmerkung: Unterstreichungen kennzeichnen die besten Resultate je Kriterium

Die größten Verzerrungen treten unter den negativen hypergeometrischen Verteilungen mit Effektgrößen von 1.6 auf, wobei hier sämtliche Schätzer auch bei $n = 60$ noch zu deutlicher Unterschätzung des Parameters neigen (Abbildung 27). Als generelles Maß kann hier $\hat{d}'_{c1,R}$ empfohlen werden, sofern aber eine Überschätzung vermieden werden soll, bieten sich auch die Bootstrap-Schätzer $\hat{d}'^{Boot}_{c1,R}$ und $\hat{d}'^{Boot}_{c2,R}$ an.

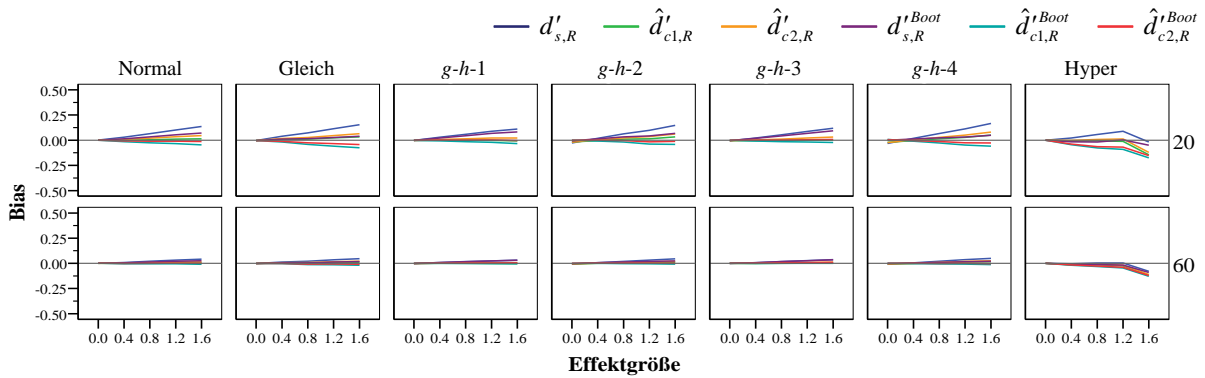


Abbildung 27. Mittlere Verzerrungen der Schätzer für d'_R getrennt nach Stichprobengröße

2.2.5.1.6 Schätzer für $d_R^{S_1}$ und $d_R^{S_2}$

Für die Schätzung von $d_R^{S_k}$ bietet sich insgesamt der t -korrigierte Schätzer $\hat{d}_R^{S_k}$ an, der ähnliche B_{MC}^2 -Werte wie $\hat{d}_R^{S_k,Boot}$ hat, jedoch durchgängig geringere quadratische Fehler aufweist (Tabelle 34).

Tabelle 34. $\overline{B_{MC}^2}$ und $\overline{MSE_{MC}}$ der Schätzer für $d_R^{S_1}$ und $d_R^{S_2}$

	$n = 20$		$n = 60$		Gesamt	
	$\overline{B_{MC}^2}$	$\overline{MSE_{MC}}$	$\overline{B_{MC}^2}$	$\overline{MSE_{MC}}$	$\overline{B_{MC}^2}$	$\overline{MSE_{MC}}$
$d_{s,R}^{S_1}$	0.012	0.209	0.001	0.047	0.006	0.128
$\hat{d}_R^{S_1}$	0.002	0.172	0.000	0.044	<u>0.001</u>	<u>0.108</u>
$d_{s,R}^{S_1,Boot}$	0.001	0.222	0.001	0.051	<u>0.001</u>	0.136
$\hat{d}_R^{S_1,Boot}$	0.004	0.195	0.001	0.049	0.002	0.122
$d_{s,R}^{S_2}$	0.019	0.271	0.003	0.074	0.011	0.173
$\hat{d}_R^{S_2}$	0.005	0.223	0.001	0.070	0.003	<u>0.146</u>
$d_{s,R}^{S_2,Boot}$	0.004	0.309	0.001	0.093	0.003	0.201
$\hat{d}_R^{S_2,Boot}$	0.004	0.268	0.001	0.089	<u>0.002</u>	0.178

Anmerkung: Unterstreichungen kennzeichnen die besten Resultate je Kriterium

Dem Bootstrap wäre hier wiederum der Vorzug zu geben, sofern eine Überschätzung des Parameters vermieden werden soll. Denn während der t -korrigierte Schätzer positive Verzerrungen aufweist, sind diese beim Bootstrap in aller Regel negativ (Abbildung 28).

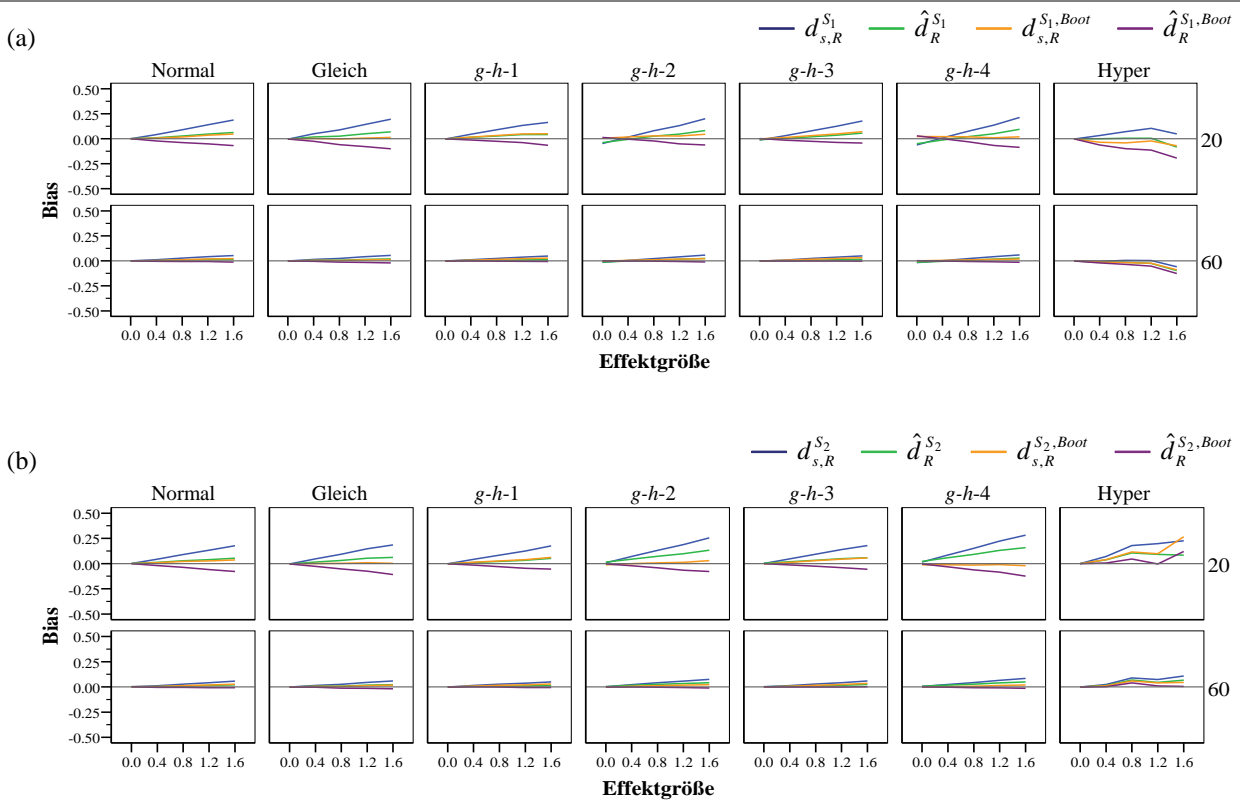


Abbildung 28. Mittlere Verzerrungen der Schätzer für $d_{s,R}^{S_1}$ und $d_{s,R}^{S_2}$ getrennt nach Stichprobengröße

2.2.5.2 Intervallschätzer

2.2.5.2.1 Schätzer für d_D

Von den für die Schätzung von d_D ausgewählten Statistiken schneiden über alle 420 Simulationsbedingungen hinweg betrachtet das nichtzentrale Intervall $NC(d_{D,s})$ mit dem größten Anteil von Intervallen mit einer beobachteten Überdeckungswahrscheinlichkeit von $\geq .936$ ($pCP_{.936}^{1.000}$) sowie die beiden BCA-Intervalle mit den geringsten absoluten Abweichungen vom nominalen Niveau ($|\overline{CP} - .95|$) ab (vgl. Tabelle 35).

Tabelle 35. Simulierte Überdeckungswahrscheinlichkeiten und Intervalllängen der Schätzer für d_D über alle Bedingungen

	Überdeckung						Länge	
	$pCP_{.936}^{.966}$	$pCP_{.936}^{1.000}$	$ \overline{CP} - .95 $	Mittel	Min	Max	Mittel	Max
Perc ($d_{D,s}$)	.243	.243	.064	.886	.548	.951	1.021	2.252
BCA ($d_{D,s}$)	<u>.562</u>	.571	.033	.921	.686	.972	0.977	2.033
Perc-t ($d_{D,s}$)	.414	.483	.039	.916	.681	.985	0.926	1.831
Perc (\hat{d}_D)	.279	.279	.055	.895	.599	.954	0.990	2.162
BCA (\hat{d}_D)	.529	.538	<u>.031</u>	.922	.721	.975	0.947	1.951
Perc-t (\hat{d}_D)	.350	.405	.041	.914	.696	.988	0.899	1.758
NC ($d_{D,s}$)	.529	<u>.681</u>	.040	.923	.501	1.000	0.942	1.751

Anmerkung: Unterstreichungen kennzeichnen die besten Resultate je Kriterium

Da die Unterschiede zwischen den beiden BCA-Verfahren gering sind, wird der Übersichtlichkeit halber in Abbildung 29 nur $\text{BCA}(\hat{d}_D)$ dem nichtzentralen Intervall für sämtliche Bedingungen mit $n = 20$ gegenübergestellt. Die anderen Verfahren, deren Überdeckungsleistungen zum Teil deutlich schlechter ausfallen, werden hier nicht weiter berücksichtigt, detaillierte Ergebnisse finden sich im Anhang (Tabelle D-17 und Tabelle D-18). Unter der Normalverteilung hält das nichtzentrale Intervall $\text{NC}(d_{D,s})$ über alle Bedingungen hinweg das nominale Niveau ein. Die BCA-Intervalle hingegen unterschreiten mit steigender Effektgröße die nominale Überdeckungswahrscheinlichkeit zunehmend. Dieses Verhalten zeigen alle Verfahren bei den g - h -Verteilungen unter Varianzhomogenität und insbesondere mit Zunahme der Korrelation zwischen den Gruppen. Unter den schiefen g - h -Verteilungen mit weniger starken und kurzen Tails (g - h -2 und g - h -4) wird das NC-Intervall bei Varianzheterogenität mit steigendem Effekt eher konservativ. Die BCA-Intervalle hingegen tendieren auch dann zur Unterschreitung des Konfidenzniveaus.

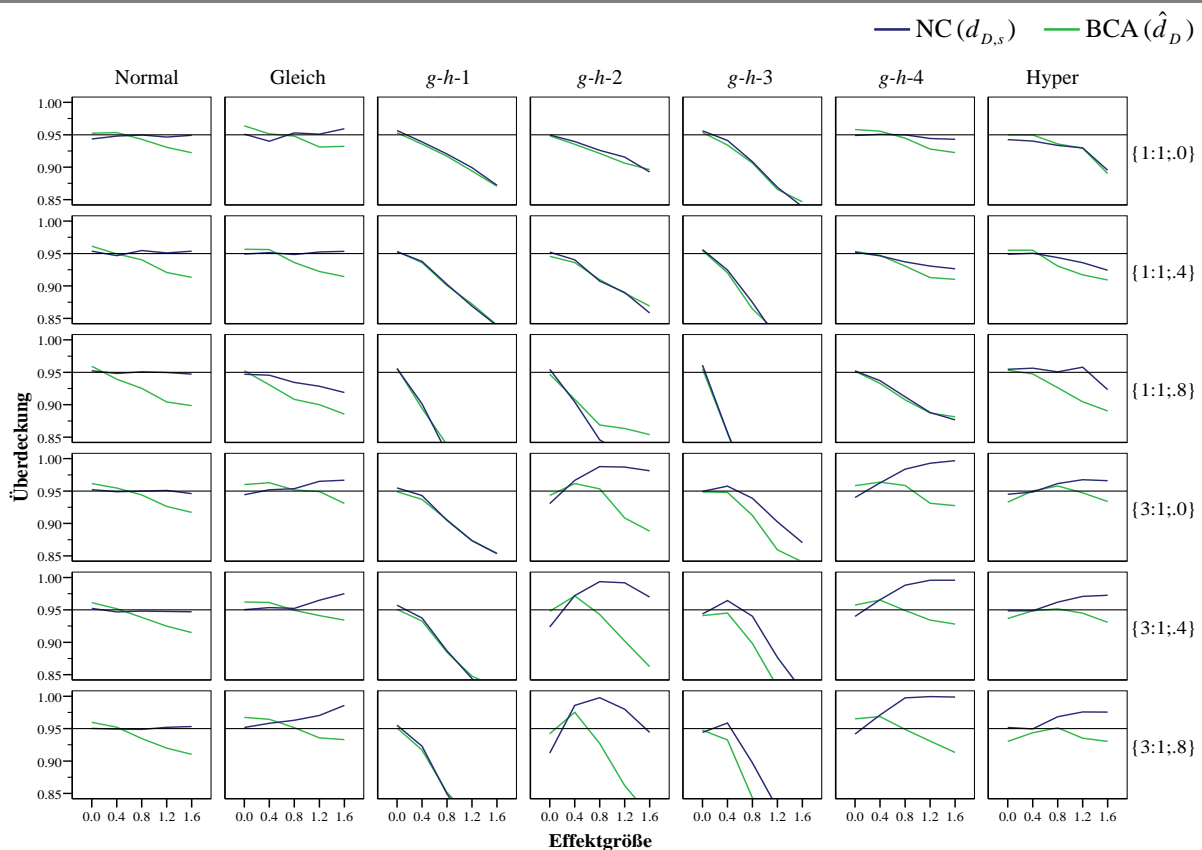


Abbildung 29. Überdeckungswahrscheinlichkeiten ausgewählter Schätzer für d_D bei $n = 20$

Im Hinblick auf die Überdeckung schneidet das nichtzentrale Intervall bei $n = 20$ insgesamt besser ab als die BCA-Intervalle. Wie die detaillierte Auswertung getrennt nach Verteilung und Stichprobengröße (Tabelle 36) verdeutlicht, kehrt sich dieses Verhältnis mit steigendem n zugunsten der BCA-Intervalle um. Bei $n = 60$ haben die BCA-Intervalle vor allem im Hinblick auf die absolute Abweichung vom nominalen Niveau zumeist – außer unter der Normalverteilung – bessere Eigenschaften als das NC-Intervall.

Tabelle 36. Simulierte Überdeckungswahrscheinlichkeiten und Intervallängen ausgewählter Schätzer für d_D getrennt nach Verteilung und Stichprobengröße

	n		Überdeckung					Länge		
			$pCP_{.936}^{.966}$	$pCP_{.936}^{1.000}$	$ CP - .95 $	Mittel	Min	Max	Mittel	Max
Normal	20	NC($d_{D,s}$)	1.000	1.000	.002	.950	.944	.955	1.192	1.643
		BCA($d_{D,s}$)	.533	.533	.017	.937	.903	.962	1.225	1.643
		BCA(\hat{d}_D)	.533	.533	.018	.936	.899	.962	1.176	1.577
	60	NC($d_{D,s}$)	1.000	1.000	.002	.949	.945	.954	0.676	0.926
		BCA($d_{D,s}$)	.933	.933	.007	.946	.931	.959	0.681	0.925
		BCA(\hat{d}_D)	.767	.767	.007	.945	.926	.959	0.672	0.913
Gleich	20	NC($d_{D,s}$)	.767	.900	.009	.953	.919	.986	1.189	1.634
		BCA($d_{D,s}$)	.700	.733	.014	.944	.889	.967	1.212	1.584
		BCA(\hat{d}_D)	.567	.600	.016	.941	.885	.967	1.163	1.520
	60	NC($d_{D,s}$)	.733	.900	.009	.954	.928	.985	0.675	0.924
		BCA($d_{D,s}$)	.933	.933	.004	.949	.934	.957	0.664	0.880
		BCA(\hat{d}_D)	.933	.933	.005	.947	.930	.957	0.656	0.869
$g-h-1$	20	NC($d_{D,s}$)	.333	.333	.067	.885	.757	.957	1.239	1.733
		BCA($d_{D,s}$)	.233	.233	.072	.879	.760	.956	1.382	2.002
		BCA(\hat{d}_D)	.267	.267	.062	.889	.789	.956	1.327	1.922
	60	NC($d_{D,s}$)	.233	.233	.103	.848	.645	.954	0.690	0.952
		BCA($d_{D,s}$)	.233	.233	.054	.896	.808	.956	0.824	1.237
		BCA(\hat{d}_D)	.267	.267	.050	.900	.821	.956	0.813	1.221
$g-h-2$	20	NC($d_{D,s}$)	.200	.567	.041	.934	.796	.998	1.208	1.678
		BCA($d_{D,s}$)	.300	.367	.045	.909	.784	.972	1.213	1.902
		BCA(\hat{d}_D)	.333	.400	.041	.913	.816	.975	1.164	1.826
	60	NC($d_{D,s}$)	.233	.567	.043	.932	.780	.995	0.679	0.934
		BCA($d_{D,s}$)	.433	.433	.025	.928	.859	.962	0.674	1.106
		BCA(\hat{d}_D)	.433	.433	.023	.929	.869	.962	0.665	1.092
$g-h-3$	20	NC($d_{D,s}$)	.400	.400	.076	.878	.657	.964	1.255	1.751
		BCA($d_{D,s}$)	.267	.267	.093	.858	.686	.955	1.373	2.033
		BCA(\hat{d}_D)	.267	.267	.080	.871	.721	.955	1.318	1.951
	60	NC($d_{D,s}$)	.333	.333	.129	.825	.501	.961	0.696	0.959
		BCA($d_{D,s}$)	.233	.233	.082	.870	.731	.957	0.823	1.291
		BCA(\hat{d}_D)	.267	.267	.075	.876	.746	.957	0.813	1.275
$g-h-4$	20	NC($d_{D,s}$)	.500	.833	.025	.956	.877	.999	1.190	1.644
		BCA($d_{D,s}$)	.600	.633	.018	.940	.878	.968	1.131	1.701
		BCA(\hat{d}_D)	.467	.500	.020	.937	.881	.969	1.086	1.633
	60	NC($d_{D,s}$)	.433	.833	.025	.958	.880	1.000	0.676	0.926
		BCA($d_{D,s}$)	.900	.900	.006	.947	.926	.957	0.614	0.940
		BCA(\hat{d}_D)	.867	.867	.007	.946	.923	.957	0.606	0.928
Hyper	20	NC($d_{D,s}$)	.567	.800	.013	.950	.895	.976	1.151	1.635
		BCA($d_{D,s}$)	.633	.633	.015	.938	.888	.963	1.180	1.910
		BCA(\hat{d}_D)	.500	.500	.016	.936	.890	.958	1.133	1.833
	60	NC($d_{D,s}$)	.667	.833	.014	.951	.880	.979	0.676	0.931
		BCA($d_{D,s}$)	.933	.933	.005	.948	.926	.956	0.679	1.149
		BCA(\hat{d}_D)	.933	.933	.006	.946	.924	.955	0.671	1.135

2.2.5.2.2 Schätzer für d'

In Tabelle 37 sind die Eigenschaften aller Intervallschätzer für d' aufgeführt. In dieser Gesamtanalyse zeigt sich bereits, daß das Perzentil- und das Perzentil- t -Bootstrap auch für die Intervallschätzung von d' weniger gut geeignet sind. Die relativen Überdeckungshäufigkeiten ($pCP_{.936}^{1.000}$ -Werte zwischen .307 und .524) und absoluten Abweichungen vom nominalen Niveau ($|CP - .95|$ von .036 bis .053) deuten auf ein schlechteres Überdeckungsverhalten im Vergleich zu den BCA-Intervallen sowie auch zu dem nichtzentralen Intervall $NC_1(d'_s)$ hin. Das insgesamt sehr ähnliche Verhalten der drei BCA-Intervalle sei im folgenden stellvertretend durch das Verfahren $BCA(\hat{d}'_{c1})$ illustriert, da dieses die geringste absolute Abweichung vom nominalen Niveau sowie die kürzesten Intervalle aufweist.

Tabelle 37. Simulierte Überdeckungswahrscheinlichkeiten und Intervalllängen der Schätzer für d' über alle Bedingungen

	Überdeckung						Länge	
	$pCP_{.936}^{.966}$	$pCP_{.936}^{1.000}$	$ CP - .95 $	Mittel	Min	Max	Mittel	Max
Perc (d'_s)	.307	.307	.053	.897	.625	.951	0.963	2.052
BCA (d'_s)	<u>.607</u>	.610	.028	.924	.746	.967	0.934	1.900
Perc- t_1 (d'_s)	.431	.488	.039	.916	.751	.982	0.899	1.738
Perc- t_2 (d'_s)	.445	.524	.036	.920	.766	.985	0.908	1.787
Perc (\hat{d}'_{c1})	.355	.355	.045	.905	.670	.953	0.933	1.970
BCA (\hat{d}'_{c1})	.548	.552	<u>.026</u>	.925	.780	.968	0.906	1.824
Perc- t (\hat{d}'_{c1})	.374	.421	.040	.914	.764	.983	0.873	1.669
NC_1 (d'_s)	.514	<u>.729</u>	.035	.933	.601	1.000	0.938	1.738
Perc (\hat{d}'_{c2})	.350	.350	.046	.904	.666	.952	0.942	1.998
BCA (\hat{d}'_{c2})	.564	.569	<u>.026</u>	.925	.775	.967	0.914	1.847
Perc- t (\hat{d}'_{c2})	.402	.464	.038	.918	.775	.985	0.889	1.738
NC_2 (d'_s)	.512	.681	.041	.922	.554	.999	0.905	1.648

Anmerkung: Unterstreichungen kennzeichnen die besten Resultate je Kriterium

Abbildung 30 zeigt die beobachteten Überdeckungswahrscheinlichkeiten von $NC_1(d'_s)$ und von $BCA(\hat{d}'_{c1})$ für alle Situationen mit $n = 20$, wobei sich insgesamt die gleiche Ergebnisstruktur wie bei der Schätzung von d_D ergibt (vgl. Abbildung 29). Das nichtzentrale Intervall hält unter der Normalverteilung das nominale Niveau in aller Regel ein und neigt unter den Verteilungen mit kurzen Tails bei größeren Parameterwerten zu Überdeckungswahrscheinlichkeiten, die oberhalb des nominalen Wertes liegen. Das BCA-Verfahren tendiert unter allen Verteilungen bei steigenden Parameterwerten zu einer Unterschreitung des Konfidenzniveaus. Auch hier ist unter den Verteilungen mit stärkeren Tails ($g-h-1$, $g-h-2$ und $g-h-3$) und vor allem bei großen Parameterwerten keines der Verfahren zur Schätzung von d' geeignet.

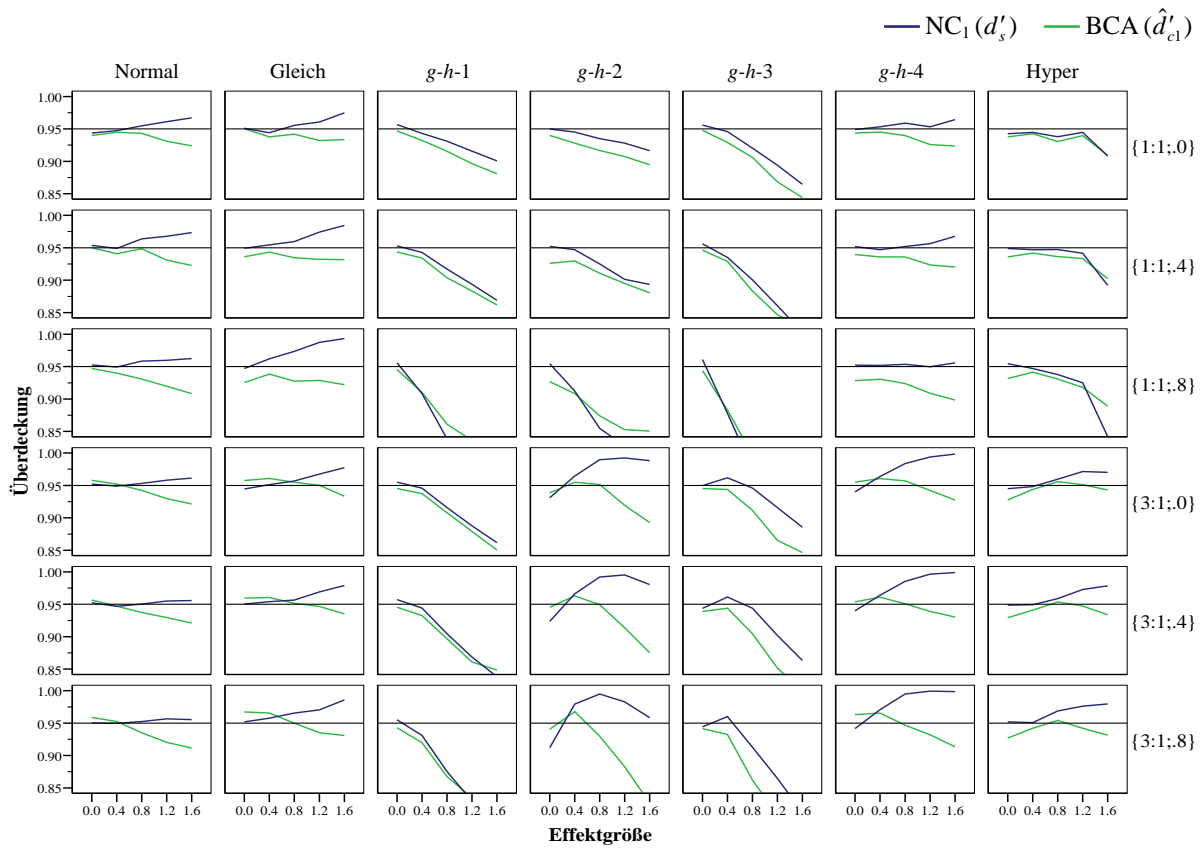


Abbildung 30. Überdeckungswahrscheinlichkeiten ausgewählter Schätzer für d' bei $n = 20$

Ebenfalls analog zu den Ergebnissen zur Schätzung von d_D führt auch hier bei $n = 20$ das nichtzentrale Intervall tendentiell zu besserer Überdeckung im Vergleich mit dem BCA-Intervall (Tabelle 38). Während aber der Übergang zu $n = 60$ durchgängig zu günstigeren Kennwerten für das BCA-Verfahren führt, bleiben die Eigenschaften des nichtzentralen Verfahrens unverändert oder verschlechtern sich sogar (siehe hierzu in Tabelle 38 z.B. $|\overline{CP} - .95|$ -Werte unter den $g-h$ -Verteilungen).

Tabelle 38. Simulierte Überdeckungswahrscheinlichkeiten und Intervallängen ausgewählter Schätzer für d' getrennt nach Verteilung und Stichprobengröße

	n		Überdeckung					Länge		
			$pCP_{.936}^{.966}$	$pCP_{.936}^{1.000}$	$ \overline{CP} - .95 $	Mittel	Min	Max	Mittel	Max
Normal	20	$NC_1(d'_s)$.900	1.000	.006	.955	.944	.973	1.187	1.635
		$BCA(\hat{d}'_{cl})$.533	.533	.015	.936	.908	.959	1.121	1.497
	60	$NC_1(d'_s)$.900	1.000	.007	.955	.946	.973	0.675	0.925
		$BCA(\hat{d}'_{cl})$.933	.933	.006	.945	.935	.954	0.649	0.873
Gleich	20	$NC_1(d'_s)$.600	1.000	.015	.964	.944	.993	1.179	1.622
		$BCA(\hat{d}'_{cl})$.533	.567	.013	.942	.922	.967	1.079	1.399
	60	$NC_1(d'_s)$.567	1.000	.016	.965	.946	.995	0.674	0.922
		$BCA(\hat{d}'_{cl})$	1.000	1.000	.003	.948	.941	.955	0.613	0.794
$g-h-1$	20	$NC_1(d'_s)$.333	.333	.053	.899	.784	.957	1.222	1.725
		$BCA(\hat{d}'_{cl})$.233	.233	.054	.896	.823	.947	1.261	1.824
	60	$NC_1(d'_s)$.233	.233	.085	.866	.694	.954	0.684	0.951
		$BCA(\hat{d}'_{cl})$.233	.233	.042	.908	.858	.949	0.783	1.178
$g-h-2$	20	$NC_1(d'_s)$.233	.567	.036	.940	.815	.995	1.203	1.656
		$BCA(\hat{d}'_{cl})$.267	.300	.039	.913	.827	.968	1.106	1.649
	60	$NC_1(d'_s)$.233	.600	.039	.937	.781	.995	0.678	0.929
		$BCA(\hat{d}'_{cl})$.433	.433	.024	.928	.868	.960	0.643	1.011
$g-h-3$	20	$NC_1(d'_s)$.400	.400	.057	.897	.709	.962	1.230	1.738
		$BCA(\hat{d}'_{cl})$.267	.267	.069	.881	.780	.948	1.243	1.758
	60	$NC_1(d'_s)$.333	.333	.101	.853	.601	.963	0.687	0.958
		$BCA(\hat{d}'_{cl})$.267	.267	.060	.891	.819	.953	0.775	1.175
$g-h-4$	20	$NC_1(d'_s)$.633	1.000	.018	.966	.940	.999	1.182	1.616
		$BCA(\hat{d}'_{cl})$.500	.500	.017	.937	.898	.966	1.013	1.417
	60	$NC_1(d'_s)$.533	1.000	.021	.970	.946	1.000	0.674	0.923
		$BCA(\hat{d}'_{cl})$.967	.967	.004	.947	.934	.956	0.570	0.828
Hyper	20	$NC_1(d'_s)$.633	.867	.016	.946	.842	.979	1.184	1.618
		$BCA(\hat{d}'_{cl})$.533	.533	.016	.935	.889	.956	1.156	1.735
	60	$NC_1(d'_s)$.667	.867	.017	.949	.852	.983	0.675	0.926
		$BCA(\hat{d}'_{cl})$.967	.967	.004	.947	.933	.954	0.670	1.076

2.2.5.2.3 Schätzer für d^{S_1} und d^{S_2}

Wie Tabelle 39 deutlich macht, ergeben insgesamt wiederum die BCA-Verfahren die robustesten Intervalle, wobei die Variante mit korrigiertem Schätzer $BCA(\hat{d}^{S_k})$ die geringste absolute Abweichung vom nominalen Niveau aufweist. Dieses Verfahren sei im folgenden $NC(d^{S_k})$ gegenübergestellt.

Tabelle 39. Simulierte Überdeckungswahrscheinlichkeiten und Intervallängen der Schätzer für d^{S_1} und d^{S_2} über alle Bedingungen

	Überdeckung						Länge	
	$pCP_{.936}^{.966}$	$pCP_{.936}^{1.000}$	$ CP - .95 $	Mittel	Min	Max	Mittel	Max
Perc ($d_s^{S_1}$)	.324	.324	.055	.895	.620	.952	1.008	2.436
BCA ($d_s^{S_1}$)	<u>.626</u>	.636	.029	.924	.744	.968	0.967	2.168
Perc-t ($d_s^{S_1}$)	.481	.560	.037	.919	.753	.985	0.915	1.959
Perc ($\hat{d}_s^{S_1}$)	.360	.360	.047	.903	.664	.955	0.977	2.338
BCA ($\hat{d}_s^{S_1}$)	.583	.593	<u>.027</u>	.925	.770	.971	0.938	2.081
Perc-t ($\hat{d}_s^{S_1}$)	.417	.481	.039	.916	.764	.986	0.887	1.881
NC ($d_s^{S_1}$)	.505	<u>.762</u>	.037	.932	.586	.999	0.953	1.854
Perc ($d_s^{S_2}$)	.271	.271	.054	.896	.618	.950	1.182	3.621
BCA ($d_s^{S_2}$)	<u>.529</u>	<u>.529</u>	.032	.919	.742	.964	1.115	3.248
Perc-t ($d_s^{S_2}$)	.443	.448	.041	.911	.756	.966	1.062	2.547
Perc ($\hat{d}_s^{S_2}$)	.314	.314	.047	.903	.662	.950	1.145	3.476
BCA ($\hat{d}_s^{S_2}$)	.490	.490	<u>.031</u>	.920	.773	.964	1.081	3.118
Perc-t ($\hat{d}_s^{S_2}$)	.398	.398	.042	.909	.767	.965	1.031	2.445
NC ($d_s^{S_2}$)	.460	.521	.056	.899	.544	.992	0.972	1.923

Anmerkung: Unterstreichungen kennzeichnen die besten Resultate je Kriterium

Wie Abbildung 31 und Abbildung 32 zeigen hat die Korrelation zwischen den Gruppen keinen wesentlichen Einfluß auf die Überdeckungsleistung der Verfahren. Das nichtzentrale Verfahren hält unter der Normalverteilung durchgängig das nominale Niveau ein und bietet Vorteile gegenüber den BCA-Verfahren, die auch hier mit steigendem Parameter die nominale Überdeckungswahrscheinlichkeit unterschreiten. Dieses Verhalten zeigen die BCA-Verfahren auch unter allen anderen Verteilungen, wobei das Ausmaß der Abweichung vom nominalen Niveau vor allem bei starken Tails groß ist. Unter den Modellen mit kurzen Tails überschreitet $NC(d_s^{S_1})$ tendentiell die nominale Wahrscheinlichkeit, während für $NC(d_s^{S_2})$ unter $g-h-4$ ebenfalls deutliche Unterschreitungen des Niveaus zu beobachten sind. Wie im unabhängigen Fall zeigt sich auch hier, daß die Schätzung von d^{S_1} unter den hier betrachteten Bedingungen, nämlich positive Effekte bei rechtsschiefen Verteilungen, etwas besser gelingt als von d^{S_2} .

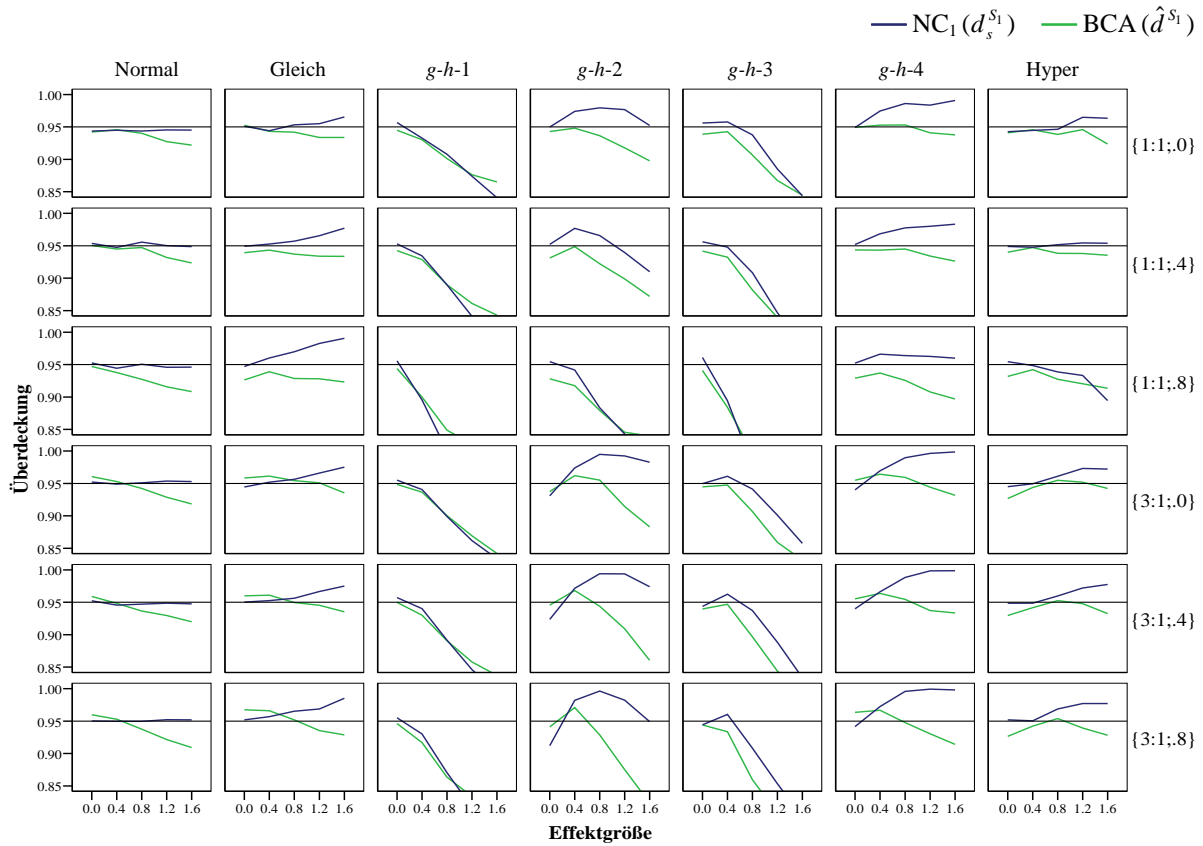


Abbildung 31. Überdeckungswahrscheinlichkeiten ausgewählter Schätzer für d^{S_1} bei $n = 20$

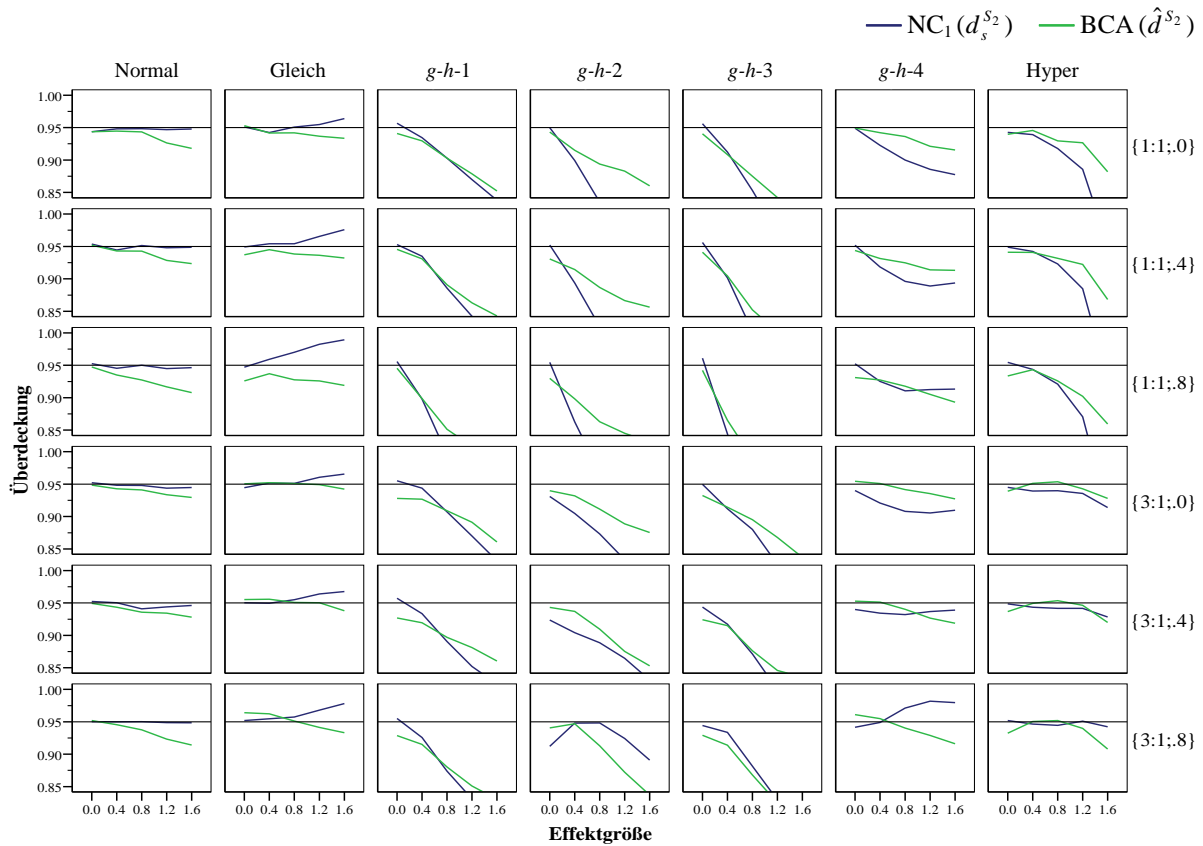


Abbildung 32. Überdeckungswahrscheinlichkeiten ausgewählter Schätzer für d^{S_2} bei $n = 20$

Tabelle 40 und Tabelle 41 geben die Charakteristika der ausgewählten Verfahren getrennt nach Verteilung und Stichprobengröße wieder.

Tabelle 40. Simulierte Überdeckungswahrscheinlichkeiten und Intervallängen ausgewählter Schätzer für d^{S_1} getrennt nach Verteilung und Stichprobengröße

	n		Überdeckung					Länge		
			$pCP_{.936}^{.966}$	$pCP_{.936}^{1.000}$	$ CP - .95 $	Mittel	Min	Max	Mittel	Max
Normal	20	$NC_1(d_s^{S_1})$	1.000	1.000	.003	.949	.944	.956	1.199	1.648
		$BCA(d_s^{S_1})$.567	.567	.015	.938	.915	.960	1.235	1.648
		$BCA(\hat{d}_s^{S_1})$.567	.567	.016	.936	.908	.960	1.185	1.582
	60	$NC_1(d_s^{S_1})$	1.000	1.000	.002	.950	.945	.956	0.677	0.929
		$BCA(d_s^{S_1})$	1.000	1.000	.006	.946	.936	.955	0.682	0.927
		$BCA(\hat{d}_s^{S_1})$.933	.933	.006	.945	.931	.955	0.674	0.915
Gleich	20	$NC_1(d_s^{S_1})$.667	1.000	.012	.961	.944	.991	1.184	1.627
		$BCA(d_s^{S_1})$.900	.933	.008	.948	.926	.967	1.166	1.509
		$BCA(\hat{d}_s^{S_1})$.567	.600	.012	.943	.923	.967	1.120	1.449
	60	$NC_1(d_s^{S_1})$.633	1.000	.014	.963	.946	.992	0.675	0.923
		$BCA(d_s^{S_1})$	1.000	1.000	.003	.949	.940	.956	0.635	0.823
		$BCA(\hat{d}_s^{S_1})$	1.000	1.000	.004	.948	.940	.956	0.627	0.812
$g-h-1$	20	$NC_1(d_s^{S_1})$.267	.267	.071	.881	.749	.957	1.267	1.808
		$BCA(d_s^{S_1})$.200	.200	.071	.879	.786	.950	1.457	2.168
		$BCA(\hat{d}_s^{S_1})$.233	.233	.061	.889	.816	.950	1.399	2.081
	60	$NC_1(d_s^{S_1})$.200	.200	.107	.844	.660	.954	0.696	0.970
		$BCA(d_s^{S_1})$.233	.233	.052	.898	.841	.950	0.852	1.296
		$BCA(\hat{d}_s^{S_1})$.233	.233	.048	.902	.850	.950	0.841	1.279
$g-h-2$	20	$NC_1(d_s^{S_1})$.267	.767	.033	.952	.808	.996	1.235	1.743
		$BCA(d_s^{S_1})$.333	.400	.043	.910	.795	.968	1.194	1.700
		$BCA(\hat{d}_s^{S_1})$.333	.400	.039	.915	.823	.971	1.146	1.632
	60	$NC_1(d_s^{S_1})$.267	.733	.034	.948	.787	.996	0.685	0.947
		$BCA(d_s^{S_1})$.433	.433	.027	.925	.859	.963	0.632	0.975
		$BCA(\hat{d}_s^{S_1})$.467	.467	.026	.926	.865	.962	0.624	0.963
$g-h-3$	20	$NC_1(d_s^{S_1})$.467	.467	.062	.892	.696	.962	1.282	1.854
		$BCA(d_s^{S_1})$.300	.300	.082	.868	.744	.947	1.415	2.136
		$BCA(\hat{d}_s^{S_1})$.300	.300	.071	.879	.770	.947	1.358	2.051
	60	$NC_1(d_s^{S_1})$.367	.367	.107	.847	.586	.964	0.703	0.997
		$BCA(d_s^{S_1})$.300	.300	.069	.881	.799	.955	0.817	1.268
		$BCA(\hat{d}_s^{S_1})$.300	.300	.064	.886	.809	.955	0.807	1.252
$g-h-4$	20	$NC_1(d_s^{S_1})$.300	1.000	.027	.975	.940	.999	1.193	1.642
		$BCA(d_s^{S_1})$.767	.800	.010	.946	.905	.966	1.050	1.458
		$BCA(\hat{d}_s^{S_1})$.633	.667	.014	.941	.897	.967	1.008	1.400
	60	$NC_1(d_s^{S_1})$.200	1.000	.029	.978	.946	.999	0.675	0.928
		$BCA(d_s^{S_1})$	1.000	1.000	.003	.950	.939	.956	0.549	0.754
		$BCA(\hat{d}_s^{S_1})$.967	.967	.005	.947	.934	.956	0.542	0.744
Hyper	20	$NC_1(d_s^{S_1})$.700	.933	.012	.954	.894	.977	1.188	1.637
		$BCA(d_s^{S_1})$.733	.733	.010	.943	.923	.961	1.200	1.581
		$BCA(\hat{d}_s^{S_1})$.633	.633	.013	.938	.913	.955	1.152	1.518
	60	$NC_1(d_s^{S_1})$.733	.933	.011	.957	.914	.981	0.676	0.927
		$BCA(d_s^{S_1})$	1.000	1.000	.002	.949	.941	.957	0.659	0.873
		$BCA(\hat{d}_s^{S_1})$	1.000	1.000	.003	.948	.939	.954	0.650	0.862

Tabelle 41. Simulierte Überdeckungswahrscheinlichkeiten und Intervallängen ausgewählter Schätzer für d^{S_2} getrennt nach Verteilung und Stichprobengröße

	n		Überdeckung				Länge			
			$pCP_{.936}^{.966}$	$pCP_{.936}^{1.000}$	$ CP - .95 $	Mittel	Min	Max	Mittel	Max
Normal	20	$NC_1(d_s^{S_2})$	1.000	1.000	.003	.948	.941	.954	1.209	1.664
		$BCA(d_s^{S_2})$.667	.667	.013	.938	.909	.952	1.247	1.679
		$BCA(\hat{d}_s^{S_2})$.500	.500	.015	.935	.908	.952	1.197	1.612
	60	$NC_1(d_s^{S_2})$	1.000	1.000	.003	.949	.943	.957	0.680	0.934
		$BCA(d_s^{S_2})$.933	.933	.005	.945	.935	.953	0.686	0.935
		$BCA(\hat{d}_s^{S_2})$.867	.867	.007	.944	.932	.953	0.677	0.923
Gleich	20	$NC_1(d_s^{S_2})$.767	1.000	.010	.959	.942	.989	1.189	1.631
		$BCA(d_s^{S_2})$.967	.967	.007	.948	.926	.964	1.155	1.498
		$BCA(\hat{d}_s^{S_2})$.767	.767	.010	.943	.919	.964	1.109	1.438
	60	$NC_1(d_s^{S_2})$.600	1.000	.014	.963	.946	.992	0.676	0.928
		$BCA(d_s^{S_2})$	1.000	1.000	.003	.949	.941	.955	0.633	0.824
		$BCA(\hat{d}_s^{S_2})$	1.000	1.000	.004	.947	.940	.955	0.625	0.814
$g-h-1$	20	$NC_1(d_s^{S_2})$.233	.233	.071	.881	.751	.957	1.306	1.847
		$BCA(d_s^{S_2})$.100	.100	.067	.883	.789	.946	1.545	2.257
		$BCA(\hat{d}_s^{S_2})$.100	.100	.059	.891	.814	.946	1.483	2.167
	60	$NC_1(d_s^{S_2})$.200	.200	.108	.843	.654	.954	0.707	0.988
		$BCA(d_s^{S_2})$.100	.100	.052	.898	.838	.946	0.879	1.327
		$BCA(\hat{d}_s^{S_2})$.100	.100	.049	.901	.846	.946	0.867	1.310
$g-h-2$	20	$NC_1(d_s^{S_2})$.167	.167	.086	.864	.740	.955	1.284	1.799
		$BCA(d_s^{S_2})$.200	.200	.057	.893	.818	.949	1.627	2.670
		$BCA(\hat{d}_s^{S_2})$.200	.200	.054	.896	.831	.947	1.562	2.563
	60	$NC_1(d_s^{S_2})$.200	.200	.096	.855	.696	.959	0.695	0.960
		$BCA(d_s^{S_2})$.300	.300	.030	.920	.881	.956	0.890	1.506
		$BCA(\hat{d}_s^{S_2})$.300	.300	.029	.921	.883	.956	0.878	1.487
$g-h-3$	20	$NC_1(d_s^{S_2})$.200	.200	.106	.846	.667	.961	1.338	1.923
		$BCA(d_s^{S_2})$.100	.100	.093	.857	.742	.942	1.642	2.490
		$BCA(\hat{d}_s^{S_2})$.100	.100	.082	.868	.773	.942	1.576	2.391
	60	$NC_1(d_s^{S_2})$.200	.200	.165	.787	.544	.959	0.720	1.015
		$BCA(d_s^{S_2})$.133	.133	.079	.872	.791	.950	0.959	1.554
		$BCA(\hat{d}_s^{S_2})$.133	.133	.073	.877	.802	.950	0.947	1.534
$g-h-4$	20	$NC_1(d_s^{S_2})$.300	.400	.030	.926	.877	.982	1.213	1.674
		$BCA(d_s^{S_2})$.467	.467	.017	.934	.902	.961	1.385	2.184
		$BCA(\hat{d}_s^{S_2})$.433	.433	.019	.932	.893	.961	1.330	2.097
	60	$NC_1(d_s^{S_2})$.333	.467	.024	.935	.882	.987	0.679	0.931
		$BCA(d_s^{S_2})$	1.000	1.000	.003	.949	.938	.956	0.730	1.166
		$BCA(\hat{d}_s^{S_2})$	1.000	1.000	.004	.948	.938	.956	0.721	1.151
Hyper	20	$NC_1(d_s^{S_2})$.600	.600	.036	.914	.723	.955	1.229	1.755
		$BCA(d_s^{S_2})$.533	.533	.020	.931	.858	.955	1.449	3.248
		$BCA(\hat{d}_s^{S_2})$.533	.533	.021	.930	.859	.954	1.391	3.118
	60	$NC_1(d_s^{S_2})$.633	.633	.033	.917	.734	.954	0.683	0.949
		$BCA(d_s^{S_2})$.900	.900	.006	.946	.926	.956	0.786	1.639
		$BCA(\hat{d}_s^{S_2})$.833	.833	.006	.945	.923	.955	0.776	1.618

Das BCA-Verfahren ohne korrigierten Punktschätzer schneidet bei $n = 20$ unter Verteilungen mit kurzen Tails besser ab als das BCA-Verfahren mit Korrektur; unter Verteilungen mit stärkeren Tails ist dies umgekehrt. Bei $n = 60$ unterscheiden sich die beiden BCA-Verfahren nicht wesentlich. Zudem zeigt sich wiederum, daß die BCA-Verfahren das nominale Niveau mit steigendem n durchweg besser einhalten, was für das nichtzentrale Intervall nicht zu beobachten ist.

2.2.5.2.4 Schätzer für $d_{D,R}$

Der einleitende Überblick zu den Charakteristika der Schätzer (Tabelle 42) zeigt, daß wie im Falle unabhängiger Gruppen die simulierten Überdeckungswahrscheinlichkeiten der Perzentil- und der BCA-Bootstrap-Verfahren der nominalen Wahrscheinlichkeit insgesamt am nächsten sind, wohingegen die Perzentil- t -Verfahren mit $|\overline{CP} - .95|$ -Werten von .024 und .030 deutlich schlechter abschneiden. Da das nichtzentrale Verfahren das nominale Niveau seltener einhält als die BCA-Verfahren und für getrimmte Statistiken auch nicht den Vorteil bietet, daß es aufgrund der in der Literatur üblicherweise berichteten (KQ-)Statistiken im nachhinein berechnet werden kann, wird das NC-Verfahren im folgenden nicht näher analysiert (zu Details siehe Tabelle D-33 und Tabelle D-34).

Tabelle 42. Simulierte Überdeckungswahrscheinlichkeiten und Intervallängen der Schätzer für $d_{D,R}$ über alle Bedingungen

	Überdeckung						Länge	
	$pCP_{.936}^{.966}$	$pCP_{.936}^{1.000}$	$ \overline{CP} - .95 $	Mittel	Min	Max	Mittel	Max
Perc ($d_{D,S,R}$)	.719	.957	.011	.955	.918	.991	1.325	3.792
BCA ($d_{D,S,R}$)	<u>.788</u>	.788	<u>.010</u>	.940	.850	.961	1.308	3.631
Perc-t ($d_{D,S,R}$)	.364	.443	.024	.932	.756	.997	1.177	3.419
Perc ($\hat{d}_{D,R}$)	.702	<u>.971</u>	.011	.957	.931	.988	1.253	3.527
BCA ($\hat{d}_{D,R}$)	<u>.788</u>	.788	.011	.940	.837	.959	1.237	3.378
Perc-t ($\hat{d}_{D,R}$)	.302	.381	.030	.926	.741	.996	1.114	3.193
NC ($d_{D,S,R}$)	.643	.748	.017	.943	.719	.995	1.070	2.280

Anmerkung: Unterstreichungen kennzeichnen die besten Resultate je Kriterium

Vergleichbar mit den Ergebnissen zur Schätzung von d'_R bei unabhängigen Gruppen läßt sich auch hier feststellen, daß das Perzentil-Bootstrap tendentiell die nominale Überdeckungswahrscheinlichkeit überschreitet, während das BCA-Intervall bei $n = 20$ diese mit zunehmendem Effekt eher unterschreitet (siehe Abbildung 33). Im diskreten, negativen hypergeometrischen Modell ist das Perzentil-Bootstrap das einzige Verfahren, welches die nominale Überdeckungswahrscheinlichkeit nicht unterschreitet. Bei Verteilungen mit starken Tails neigt das Perzentil-Verfahren mit zunehmendem Effekt dazu, die nominale Wahrscheinlichkeit zu überschreiten. Bis auf varianzhomogene Bedingungen mit hoher Korrelation zwischen den Gruppen ist auch das BCA-Verfahren für stetige Verteilungen gut geeignet. Sofern aber das α -Niveau nicht überschritten werden soll, wäre generell dem Perzentil-Verfahren der Vorzug zu geben. In der weiteren Analyse werden nun das Perzentil- und das BCA-Bootstrap mit korrigierten Punktschätzern berücksichtigt, da die anderen Verfahren diesen beiden auch bei $n = 60$ in aller Regel unterlegen sind (vgl. Anhang D).

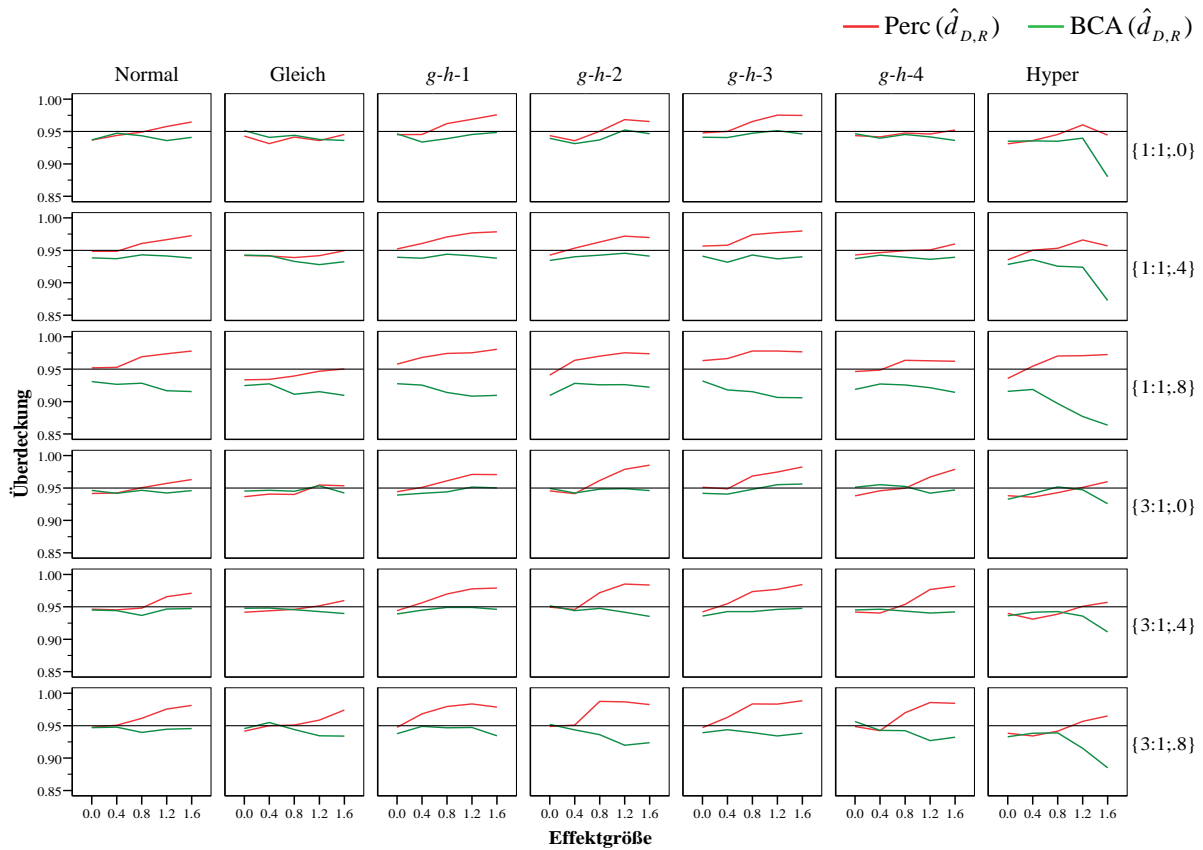


Abbildung 33. Überdeckungswahrscheinlichkeiten ausgewählter Schätzer für $d_{D,R}$ bei $n = 20$

Die Kennwerte in Tabelle 43 unterstreichen zum einen das Ergebnis der graphischen Analyse im Hinblick auf die robustere Überdeckung durch das Perzentil-Verfahren und zum anderen wird deutlich, daß die Perzentil-Intervalle auch bei $n = 20$ nicht wesentlich länger sind als die BCA-Intervalle.

Tabelle 43. Simulierte Überdeckungswahrscheinlichkeiten und Intervallängen ausgewählter Schätzer für $d_{D,R}$ getrennt nach Verteilung und Stichprobengröße

	n		Überdeckung					Länge		
			$pCP_{.936}^{.966}$	$pCP_{.936}^{1.000}$	$ \overline{CP} - .95 $	Mittel	Min	Max	Mittel	Max
Normal	20	Perc ($\hat{d}_{D,R}$)	.733	1.000	.011	.957	.937	.981	1.656	2.625
		BCA ($\hat{d}_{D,R}$)	.833	.833	.010	.940	.916	.948	1.641	2.587
	60	Perc ($\hat{d}_{D,R}$)	.967	1.000	.006	.954	.941	.967	0.832	1.195
		BCA ($\hat{d}_{D,R}$)	.933	.933	.005	.945	.929	.951	0.828	1.189
Gleich	20	Perc ($\hat{d}_{D,R}$)	.867	.900	.008	.945	.931	.974	1.934	2.726
		BCA ($\hat{d}_{D,R}$)	.667	.667	.012	.938	.910	.955	1.862	2.570
	60	Perc ($\hat{d}_{D,R}$)	1.000	1.000	.004	.950	.943	.962	0.948	1.271
		BCA ($\hat{d}_{D,R}$)	.933	.933	.005	.946	.928	.954	0.936	1.248
$g-h-1$	20	Perc ($\hat{d}_{D,R}$)	.400	1.000	.017	.966	.944	.983	1.403	2.691
		BCA ($\hat{d}_{D,R}$)	.767	.767	.011	.939	.908	.951	1.431	2.743
	60	Perc ($\hat{d}_{D,R}$)	.667	1.000	.011	.961	.946	.972	0.703	1.259
		BCA ($\hat{d}_{D,R}$)	1.000	1.000	.004	.946	.941	.953	0.707	1.262
$g-h-2$	20	Perc ($\hat{d}_{D,R}$)	.500	.967	.017	.963	.936	.987	1.451	2.346
		BCA ($\hat{d}_{D,R}$)	.667	.667	.012	.938	.910	.952	1.428	2.305
	60	Perc ($\hat{d}_{D,R}$)	.667	1.000	.012	.961	.943	.982	0.720	1.183
		BCA ($\hat{d}_{D,R}$)	1.000	1.000	.004	.946	.939	.955	0.716	1.175
$g-h-3$	20	Perc ($\hat{d}_{D,R}$)	.400	1.000	.019	.968	.942	.988	1.313	2.527
		BCA ($\hat{d}_{D,R}$)	.733	.733	.013	.938	.906	.956	1.345	2.560
	60	Perc ($\hat{d}_{D,R}$)	.500	1.000	.014	.963	.945	.977	0.656	1.162
		BCA ($\hat{d}_{D,R}$)	1.000	1.000	.005	.947	.939	.955	0.660	1.167
$g-h-4$	20	Perc ($\hat{d}_{D,R}$)	.767	1.000	.011	.956	.938	.986	1.682	2.550
		BCA ($\hat{d}_{D,R}$)	.767	.767	.012	.939	.914	.956	1.623	2.432
	60	Perc ($\hat{d}_{D,R}$)	.800	1.000	.009	.957	.941	.984	0.817	1.266
		BCA ($\hat{d}_{D,R}$)	.967	.967	.004	.946	.936	.953	0.808	1.248
Hyper	20	Perc ($\hat{d}_{D,R}$)	.667	.800	.011	.949	.931	.972	2.371	3.527
		BCA ($\hat{d}_{D,R}$)	.300	.300	.028	.922	.864	.952	2.289	3.378
	60	Perc ($\hat{d}_{D,R}$)	.900	.933	.007	.950	.931	.973	1.052	1.842
		BCA ($\hat{d}_{D,R}$)	.467	.467	.027	.924	.837	.959	1.040	1.839

2.2.5.2.5 Schätzer für d'_R

Auch für die Schätzung von d'_R bei abhängigen Stichproben besitzen insgesamt die Perzentil- und BCA-Bootstraps die (relativ) besten Eigenschaften (vgl. Tabelle 44). Da die Wahl der verschiedenen Punktschätzer keine wesentlichen Unterschiede im Hinblick auf die Überdeckung bedeutet, werden in der graphischen und tabellarischen Gegenüberstellung im folgenden die Varianten mit der Korrektur $c1$ gewählt, denn mit dieser ergeben sich insgesamt kürzere Intervalle.

Tabelle 44. Simulierte Überdeckungswahrscheinlichkeiten und Intervallängen der Schätzer für d'_R über alle Bedingungen

	Überdeckung						Länge	
	$pCP_{.936}^{.966}$	$pCP_{.936}^{1.000}$	$ CP - .95 $	Mittel	Min	Max	Mittel	Max
Perc ($d'_{s,R}$)	.788	<u>.981</u>	<u>.009</u>	.955	.931	.980	1.247	3.686
BCA ($d'_{s,R}$)	.769	.769	.010	.940	.865	.960	1.247	3.563
Perc- t_1 ($d'_{s,R}$)	.352	.419	.022	.933	.829	.997	1.150	2.978
Perc- t_2 ($d'_{s,R}$)	.450	.533	.019	.938	.829	.997	1.155	2.967
Perc ($\hat{d}'_{c1,R}$)	.745	.974	.010	.957	.926	.987	1.180	3.429
BCA ($\hat{d}'_{c1,R}$)	.800	.800	.010	.941	.845	.959	1.180	3.314
Perc- t ($\hat{d}'_{c1,R}$)	.314	.379	.026	.929	.798	.996	1.089	2.770
NC ₁ ($d'_{s,R}$)	.681	.826	.015	.950	.800	.995	1.073	2.271
Perc ($\hat{d}'_{c2,R}$)	.729	.979	.011	.957	.929	.986	1.200	3.496
BCA ($\hat{d}'_{c2,R}$)	<u>.810</u>	.810	.010	.941	.847	.959	1.202	3.376
Perc- t ($\hat{d}'_{c2,R}$)	.369	.457	.023	.934	.799	.996	1.113	2.792
NC ₂ ($d'_{s,R}$)	.548	.636	.022	.936	.780	.994	1.024	2.221

Anmerkung: Unterstreichungen kennzeichnen die besten Resultate je Kriterium

Auch im abhängigen Fall hat das Perzentil-Bootstrap bei der Schätzung von d'_R das vergleichsweise robusteste Überdeckungsverhalten. Die niedrigste simulierte Überdeckungswahrscheinlichkeit von Perc ($\hat{d}'_{c1,R}$) tritt mit .926 unter der negativen hypergeometrischen Verteilung bei einem Effekt von 1.6, $n = 20$ und $\{1:1; .8\}$ auf (vgl. Abbildung 34).

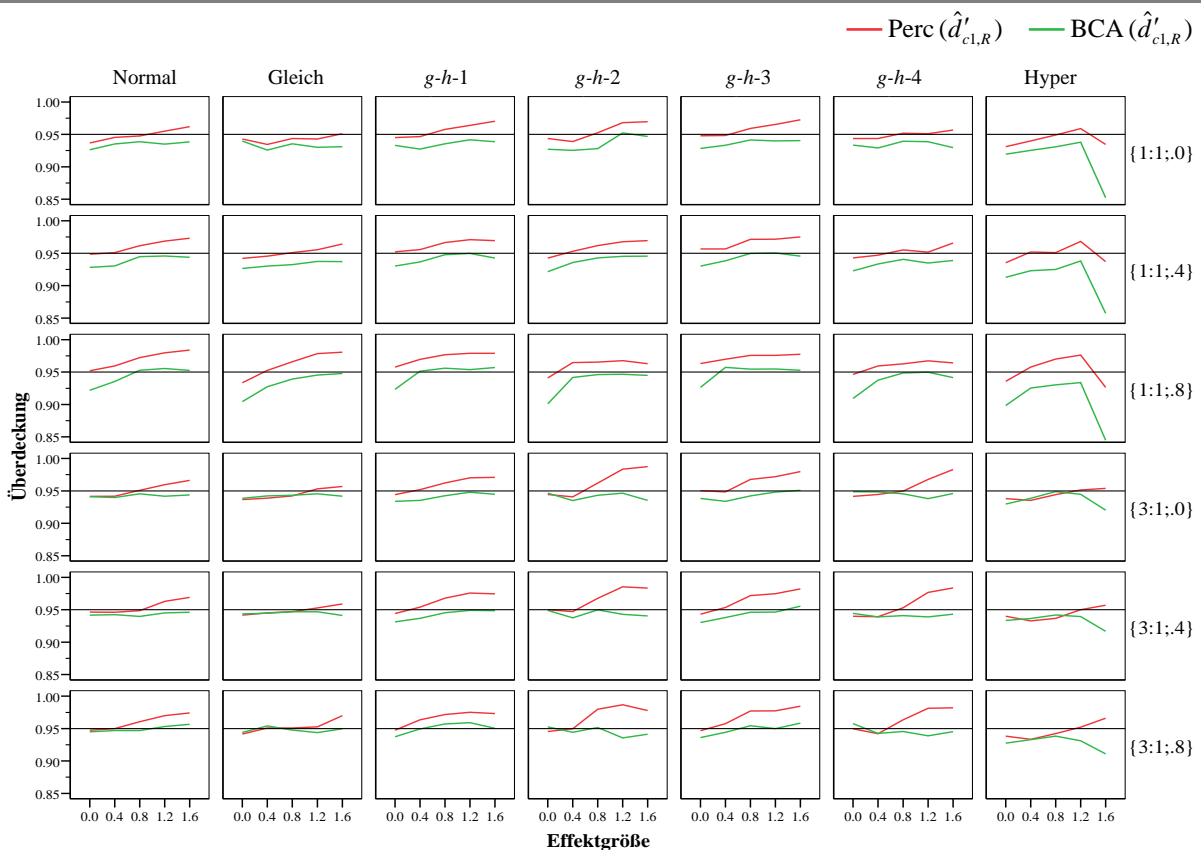


Abbildung 34. Überdeckungswahrscheinlichkeiten ausgewählter Schätzer für d'_R bei $n = 20$

In Tabelle 45 wird deutlich, daß Perc($\hat{d}'_{c1,R}$) durchweg konservativer überdeckt als BCA($\hat{d}'_{c1,R}$). Die $pCP_{.936}^{1.000}$ -Werte von Perc($\hat{d}'_{c1,R}$) liegen für alle Verteilungen und Stichprobengrößen auf oder oberhalb der Werte von BCA($\hat{d}'_{c1,R}$). Da die Perzentil-Intervalle aber nicht länger sind als die BCA-Intervalle und Perc($\hat{d}'_{c1,R}$) unter den diskreten Verteilungen deutlich besser überdeckt, kann dieses Verfahren für Stichproben ab $n = 20$ zur Schätzung von d_R empfohlen werden.

Tabelle 45. Simulierte Überdeckungswahrscheinlichkeiten und Intervalllängen ausgewählter Schätzer für d'_R getrennt nach Verteilung und Stichprobengröße

	n		Überdeckung					Länge		
			$pCP_{.936}^{.966}$	$pCP_{.936}^{1.000}$	$ CP - .95 $	Mittel	Min	Max	Mittel	Max
Normal	20	Perc($\hat{d}'_{c1,R}$)	.700	1.000	.011	.958	.937	.984	1.542	2.300
		BCA($\hat{d}'_{c1,R}$)	.767	.767	.009	.942	.922	.957	1.558	2.306
	60	Perc($\hat{d}'_{c1,R}$)	1.000	1.000	.006	.954	.941	.965	0.800	1.151
		BCA($\hat{d}'_{c1,R}$)	1.000	1.000	.004	.946	.939	.951	0.800	1.150
Gleich	20	Perc($\hat{d}'_{c1,R}$)	.833	.933	.009	.951	.933	.981	1.804	2.534
		BCA($\hat{d}'_{c1,R}$)	.700	.700	.012	.939	.904	.954	1.776	2.447
	60	Perc($\hat{d}'_{c1,R}$)	1.000	1.000	.004	.951	.944	.963	0.913	1.232
		BCA($\hat{d}'_{c1,R}$)	.967	.967	.005	.945	.933	.953	0.907	1.217
$g-h-1$	20	Perc($\hat{d}'_{c1,R}$)	.467	1.000	.015	.964	.944	.979	1.295	2.268
		BCA($\hat{d}'_{c1,R}$)	.733	.733	.009	.943	.923	.959	1.341	2.322
	60	Perc($\hat{d}'_{c1,R}$)	.800	1.000	.009	.959	.946	.970	0.670	1.095
		BCA($\hat{d}'_{c1,R}$)	1.000	1.000	.004	.946	.941	.954	0.676	1.101
$g-h-2$	20	Perc($\hat{d}'_{c1,R}$)	.567	1.000	.016	.962	.939	.987	1.357	2.296
		BCA($\hat{d}'_{c1,R}$)	.700	.700	.010	.940	.901	.953	1.355	2.195
	60	Perc($\hat{d}'_{c1,R}$)	.767	1.000	.010	.959	.944	.979	0.689	1.126
		BCA($\hat{d}'_{c1,R}$)	1.000	1.000	.005	.946	.938	.953	0.687	1.114
$g-h-3$	20	Perc($\hat{d}'_{c1,R}$)	.433	1.000	.017	.966	.943	.984	1.217	2.092
		BCA($\hat{d}'_{c1,R}$)	.800	.800	.009	.944	.926	.958	1.264	2.145
	60	Perc($\hat{d}'_{c1,R}$)	.667	1.000	.012	.961	.944	.972	0.625	0.997
		BCA($\hat{d}'_{c1,R}$)	1.000	1.000	.004	.948	.940	.955	0.630	1.002
$g-h-4$	20	Perc($\hat{d}'_{c1,R}$)	.767	1.000	.012	.957	.939	.984	1.601	2.521
		BCA($\hat{d}'_{c1,R}$)	.767	.767	.011	.940	.909	.958	1.568	2.305
	60	Perc($\hat{d}'_{c1,R}$)	.833	1.000	.009	.957	.943	.976	0.791	1.174
		BCA($\hat{d}'_{c1,R}$)	.967	.967	.004	.946	.935	.954	0.785	1.167
Hyper	20	Perc($\hat{d}'_{c1,R}$)	.633	.733	.011	.946	.926	.976	2.186	3.429
		BCA($\hat{d}'_{c1,R}$)	.300	.300	.028	.922	.845	.949	2.149	3.314
	60	Perc($\hat{d}'_{c1,R}$)	.967	.967	.005	.948	.934	.964	1.026	1.518
		BCA($\hat{d}'_{c1,R}$)	.500	.500	.023	.928	.878	.956	1.016	1.459

2.2.5.2.6 Schätzer für $d_R^{S_1}$ und $d_R^{S_2}$

Aus Tabelle 46 geht hervor, daß auch bei der Schätzung von $d_R^{S_k}$ die Intervalle auf Basis des Perzentil- sowie des BCA-Bootstraps im Hinblick auf den absoluten Fehler und die relative Anzahl robuster Intervalle deutlich bessere Eigenschaften besitzen als die anderen Verfahren.

Tabelle 46. Simulierte Überdeckungswahrscheinlichkeiten und Intervallängen der Schätzer für $d_R^{S_1}$ und $d_R^{S_2}$ über alle Bedingungen

	Überdeckung						Länge	
	$pCP_{.936}^{.966}$	$pCP_{.936}^{1.000}$	$ CP - .95 $	Mittel	Min	Max	Mittel	Max
Perc ($d_{s,R}^{S_1}$)	.719	<u>.983</u>	.011	.957	.931	.983	1.461	5.035
BCA ($d_{s,R}^{S_1}$)	.845	.845	.009	.943	.873	.962	1.459	4.757
Perc-t ($d_{s,R}^{S_1}$)	.402	.488	.021	.936	.812	.997	1.238	2.748
Perc ($\hat{d}_R^{S_1}$)	.621	.976	.013	.960	.931	.988	1.380	4.684
BCA ($\hat{d}_R^{S_1}$)	<u>.855</u>	.855	<u>.008</u>	.944	.869	.961	1.377	4.425
Perc-t ($\hat{d}_R^{S_1}$)	.324	.410	.027	.929	.781	.996	1.172	2.556
NC ($d_{s,R}^{S_1}$)	.576	.719	.019	.944	.839	.995	1.096	2.312
Perc ($d_{s,R}^{S_2}$)	.786	.979	<u>.009</u>	.955	.920	.983	1.590	4.994
BCA ($d_{s,R}^{S_2}$)	.802	.802	.012	.939	.808	.963	1.570	4.474
Perc-t ($d_{s,R}^{S_2}$)	.362	.364	.029	.921	.490	.989	1.463	6.720
Perc ($\hat{d}_R^{S_2}$)	.717	<u>.981</u>	.011	.958	.886	.987	1.503	4.646
BCA ($\hat{d}_R^{S_2}$)	<u>.817</u>	.817	.011	.940	.803	.964	1.484	4.162
Perc-t ($\hat{d}_R^{S_2}$)	.307	.310	.034	.917	.467	.989	1.385	6.251
NC ($d_{s,R}^{S_2}$)	.467	.481	.037	.916	.422	.971	1.128	2.739

Anmerkung: Unterstreichungen kennzeichnen die besten Resultate je Kriterium

Da auch die Bootstrap-Verfahren mit korrigierten Punktschätzern bei ähnlich guter Überdeckung kürzere Intervalle besitzen, wird auf die detaillierte Darstellung der Verfahren ohne Korrektur verzichtet.

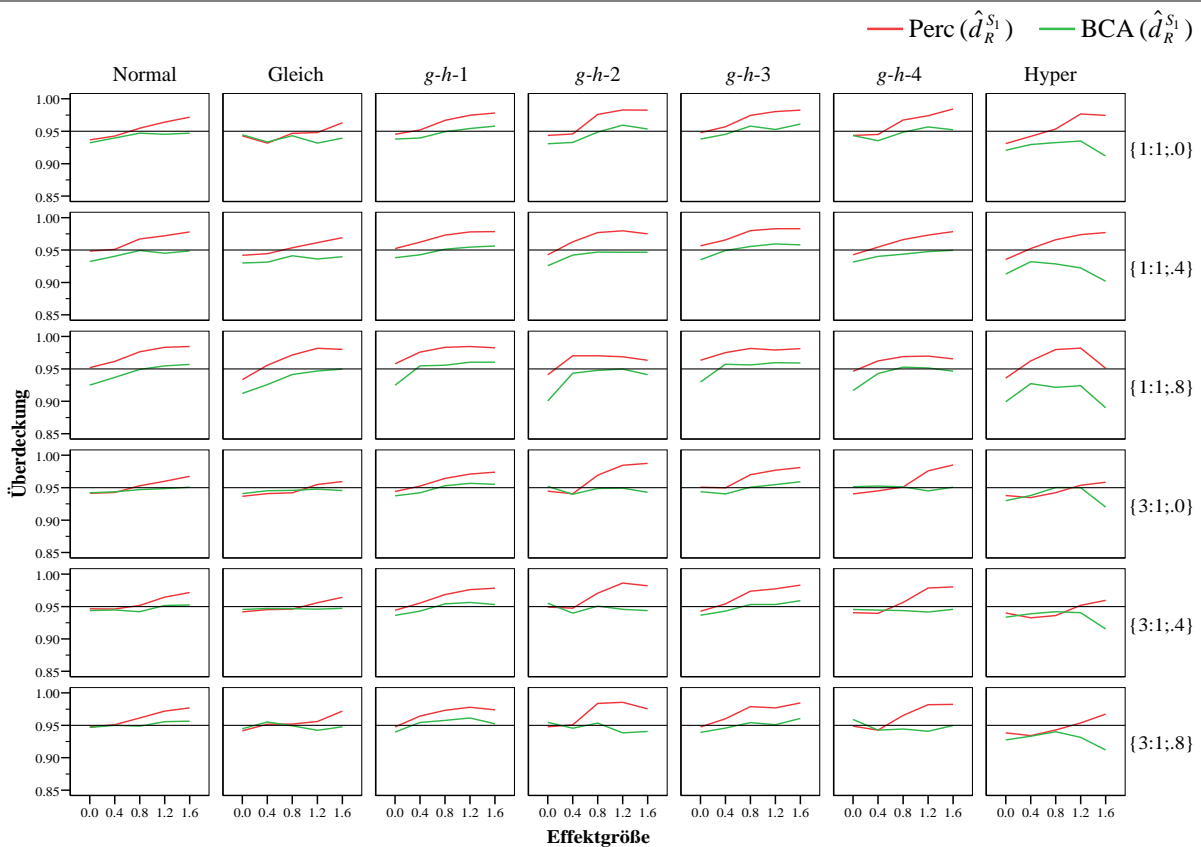


Abbildung 35. Überdeckungswahrscheinlichkeiten ausgewählter Schätzer für $d_R^{S_1}$ bei $n = 20$

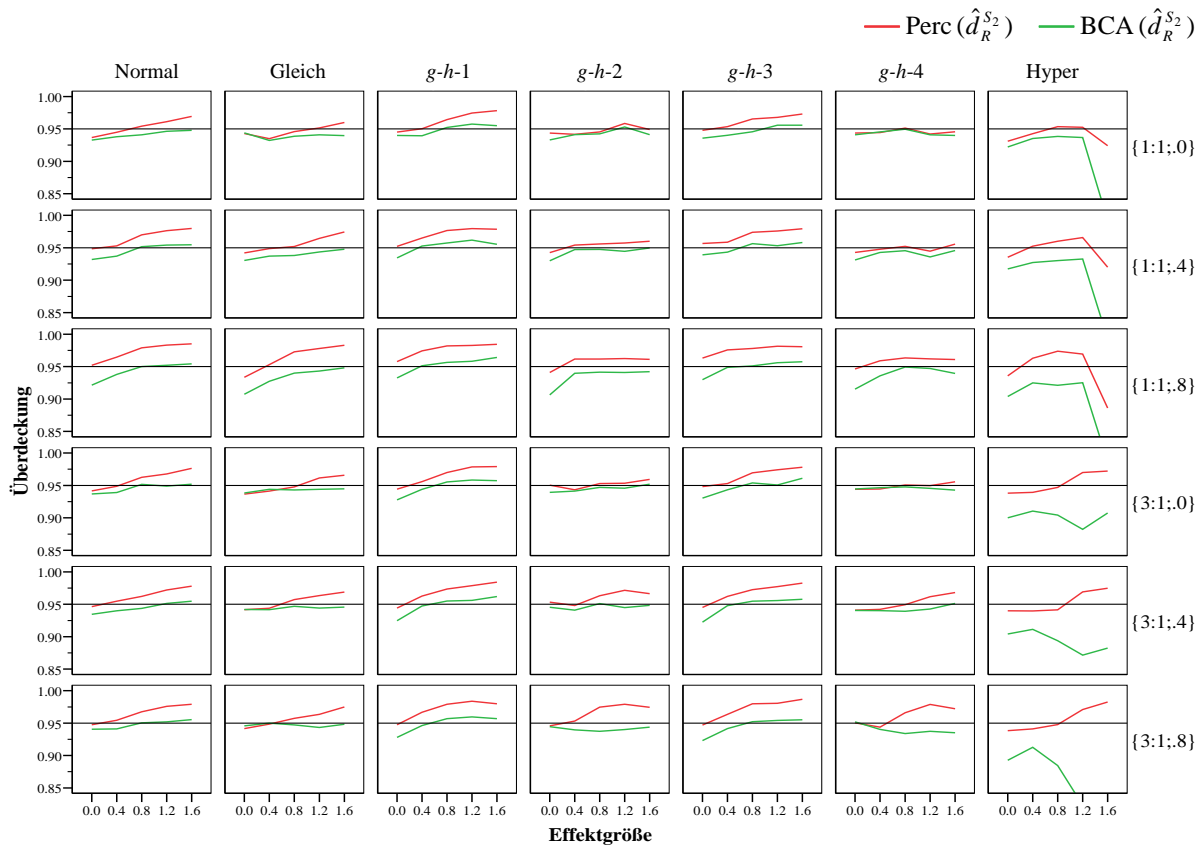


Abbildung 36. Überdeckungswahrscheinlichkeiten ausgewählter Schätzer für $d_R^{S_2}$ bei $n = 20$

Das Ergebnismuster der Perzentil- und BCA-Intervalle zur Schätzung von $d_R^{S_k}$ (vgl. Abbildung 35 und Abbildung 36 sowie Tabelle 47 und Tabelle 48) entspricht insgesamt den dargestellten Ergebnissen für die anderen robustifizierten Parameter: Das Perzentil-Verfahren zeigt für alle Bedingungen größere Überdeckungshäufigkeiten als das BCA-Verfahren, wobei die Intervalllängen der Verfahren sehr ähnlich sind. Da also das Perzentil-Verfahren die nominale Überdeckungswahrscheinlichkeit seltener unterschreitet – die extremste Unterschreitung des nominalen Niveaus tritt für Perc ($\hat{d}_R^{S_2}$) wiederum unter der negativen hypergeometrischen Verteilung mit $d_R^{S_2} = 1.6$, $n = 20$ und $\{1:1; .8\}$ auf, wobei diese mit .886 hier deutlicher ausfällt als bei den Schätzern mit gemittelter Varianz – und vergleichbare Präzision zeigt wie das BCA-Verfahren, kann Perc ($\hat{d}_R^{S_2}$) zur Schätzung von $d_R^{S_k}$ ab $n = 20$ empfohlen werden.

Tabelle 47. Simulierte Überdeckungswahrscheinlichkeiten und Intervallängen ausgewählter Schätzer für $d_R^{S_1}$ getrennt nach Verteilung und Stichprobengröße

	n		Überdeckung					Länge		
			$pCP_{.936}^{.966}$	$pCP_{.936}^{1.000}$	$ \overline{CP} - .95 $	Mittel	Min	Max	Mittel	Max
Normal	20	Perc ($\hat{d}_R^{S_1}$)	.633	1.000	.013	.960	.937	.984	1.894	2.836
		BCA ($\hat{d}_R^{S_1}$)	.900	.900	.006	.946	.925	.957	1.918	2.887
	60	Perc ($\hat{d}_R^{S_1}$)	.867	1.000	.007	.955	.941	.971	0.876	1.264
		BCA ($\hat{d}_R^{S_1}$)	1.000	1.000	.003	.947	.940	.953	0.874	1.258
Gleich	20	Perc ($\hat{d}_R^{S_1}$)	.767	.933	.011	.953	.932	.982	2.139	3.003
		BCA ($\hat{d}_R^{S_1}$)	.800	.800	.009	.941	.912	.955	2.095	2.881
	60	Perc ($\hat{d}_R^{S_1}$)	.967	1.000	.005	.953	.944	.968	0.971	1.311
		BCA ($\hat{d}_R^{S_1}$)	.967	.967	.004	.946	.935	.953	0.962	1.288
$g-h-1$	20	Perc ($\hat{d}_R^{S_1}$)	.400	1.000	.018	.967	.944	.985	1.660	2.732
		BCA ($\hat{d}_R^{S_1}$)	.967	.967	.008	.950	.925	.961	1.720	2.829
	60	Perc ($\hat{d}_R^{S_1}$)	.600	1.000	.011	.961	.946	.972	0.761	1.230
		BCA ($\hat{d}_R^{S_1}$)	1.000	1.000	.002	.949	.941	.953	0.765	1.237
$g-h-2$	20	Perc ($\hat{d}_R^{S_1}$)	.400	1.000	.019	.966	.941	.988	1.628	2.901
		BCA ($\hat{d}_R^{S_1}$)	.867	.867	.008	.944	.901	.959	1.619	2.726
	60	Perc ($\hat{d}_R^{S_1}$)	.600	1.000	.013	.962	.944	.977	0.723	1.249
		BCA ($\hat{d}_R^{S_1}$)	1.000	1.000	.004	.946	.939	.955	0.717	1.225
$g-h-3$	20	Perc ($\hat{d}_R^{S_1}$)	.367	1.000	.021	.970	.943	.985	1.538	2.659
		BCA ($\hat{d}_R^{S_1}$)	.933	.933	.008	.951	.930	.961	1.609	2.829
	60	Perc ($\hat{d}_R^{S_1}$)	.533	1.000	.014	.963	.944	.973	0.699	1.173
		BCA ($\hat{d}_R^{S_1}$)	1.000	1.000	.004	.950	.941	.957	0.703	1.182
$g-h-4$	20	Perc ($\hat{d}_R^{S_1}$)	.533	1.000	.016	.962	.939	.985	1.906	3.173
		BCA ($\hat{d}_R^{S_1}$)	.900	.900	.006	.946	.917	.959	1.850	2.807
	60	Perc ($\hat{d}_R^{S_1}$)	.667	1.000	.012	.960	.943	.979	0.806	1.199
		BCA ($\hat{d}_R^{S_1}$)	.967	.967	.003	.947	.936	.954	0.796	1.162
Hyper	20	Perc ($\hat{d}_R^{S_1}$)	.533	.800	.014	.953	.931	.982	2.667	4.684
		BCA ($\hat{d}_R^{S_1}$)	.233	.233	.024	.926	.890	.950	2.613	4.425
	60	Perc ($\hat{d}_R^{S_1}$)	.833	.933	.008	.951	.935	.970	1.053	1.445
		BCA ($\hat{d}_R^{S_1}$)	.433	.433	.027	.924	.869	.956	1.043	1.368

Tabelle 48. Simulierte Überdeckungswahrscheinlichkeiten und Intervallängen ausgewählter Schätzer für $d_R^{S_2}$ getrennt nach Verteilung und Stichprobengröße

	n		Überdeckung					Länge		
			$pCP_{.936}^{.966}$	$pCP_{.936}^{1.000}$	$ \overline{CP} - .95 $	Mittel	Min	Max	Mittel	Max
Normal	20	Perc ($\hat{d}_R^{S_2}$)	.533	1.000	.015	.963	.937	.985	1.911	2.851
		BCA ($\hat{d}_R^{S_2}$)	.867	.867	.008	.945	.921	.955	1.958	2.913
	60	Perc ($\hat{d}_R^{S_2}$)	.867	1.000	.008	.957	.941	.969	0.881	1.279
		BCA ($\hat{d}_R^{S_2}$)	1.000	1.000	.004	.946	.936	.954	0.881	1.279
Gleich	20	Perc ($\hat{d}_R^{S_2}$)	.733	.933	.012	.955	.933	.983	2.110	2.963
		BCA ($\hat{d}_R^{S_2}$)	.867	.867	.009	.941	.907	.950	2.076	2.841
	60	Perc ($\hat{d}_R^{S_2}$)	.967	1.000	.005	.953	.944	.966	0.967	1.315
		BCA ($\hat{d}_R^{S_2}$)	.967	.967	.004	.946	.935	.953	0.958	1.295
$g-h-1$	20	Perc ($\hat{d}_R^{S_2}$)	.367	1.000	.020	.969	.944	.984	1.699	2.794
		BCA ($\hat{d}_R^{S_2}$)	.833	.833	.009	.950	.925	.964	1.789	2.921
	60	Perc ($\hat{d}_R^{S_2}$)	.600	1.000	.012	.962	.946	.971	0.769	1.237
		BCA ($\hat{d}_R^{S_2}$)	1.000	1.000	.004	.949	.940	.956	0.775	1.245
$g-h-2$	20	Perc ($\hat{d}_R^{S_2}$)	.833	1.000	.009	.956	.941	.979	2.093	3.867
		BCA ($\hat{d}_R^{S_2}$)	.900	.900	.008	.942	.906	.953	2.007	3.570
	60	Perc ($\hat{d}_R^{S_2}$)	.933	1.000	.005	.954	.944	.969	0.925	1.694
		BCA ($\hat{d}_R^{S_2}$)	1.000	1.000	.003	.947	.939	.952	0.911	1.648
$g-h-3$	20	Perc ($\hat{d}_R^{S_2}$)	.400	1.000	.019	.968	.945	.987	1.679	2.871
		BCA ($\hat{d}_R^{S_2}$)	.833	.833	.008	.948	.922	.961	1.745	2.907
	60	Perc ($\hat{d}_R^{S_2}$)	.700	1.000	.012	.961	.947	.973	0.760	1.321
		BCA ($\hat{d}_R^{S_2}$)	.967	.967	.004	.948	.934	.957	0.763	1.323
$g-h-4$	20	Perc ($\hat{d}_R^{S_2}$)	.867	1.000	.008	.953	.941	.979	2.557	4.646
		BCA ($\hat{d}_R^{S_2}$)	.800	.800	.009	.941	.915	.952	2.411	4.162
	60	Perc ($\hat{d}_R^{S_2}$)	.933	1.000	.005	.953	.944	.968	1.057	1.877
		BCA ($\hat{d}_R^{S_2}$)	1.000	1.000	.003	.948	.937	.954	1.036	1.820
Hyper	20	Perc ($\hat{d}_R^{S_2}$)	.533	.800	.016	.949	.886	.982	2.263	3.542
		BCA ($\hat{d}_R^{S_2}$)	.067	.067	.055	.895	.803	.938	2.141	3.367
	60	Perc ($\hat{d}_R^{S_2}$)	.767	1.000	.010	.956	.940	.983	1.375	3.126
		BCA ($\hat{d}_R^{S_2}$)	.333	.333	.030	.920	.842	.956	1.322	2.993

2.2.5.2.7 Vergleich von KQ- und robustifizierter Schätzung

Wie bei unverbundenen Stichproben legen die Simulationsergebnisse auch bei Abhängigkeit nahe, die klassischen Parameter anhand des BCA-Bootstraps und die robustifizierten Parameter anhand des Perzentil-Bootstraps zu schätzen, wobei diese Empfehlungen sich auf die hier simulierten Modelle und Bedingungen mit minimalem $n = 20$ und nominaler Überdeckungswahrscheinlichkeit .95 beziehen. Abbildung 37, Abbildung 38 und Abbildung 39 geben die Verteilungen der beobachteten Überdeckungswahrscheinlichkeiten und Intervallängen der jeweils besten KQ- und robustifizierten Schätzer der einzelnen Parameter im Vergleich wieder. Zu beachten ist wiederum, daß bei den Gegenüberstellungen in Abbildung 39 die Ergebnisse von $BCA(\hat{d}_R^{S_1})$ und $BCA(\hat{d}_R^{S_2})$ sowie von $Perc(\hat{d}_R^{S_1})$ und $Perc(\hat{d}_R^{S_2})$ jeweils zusammengefaßt wurden.

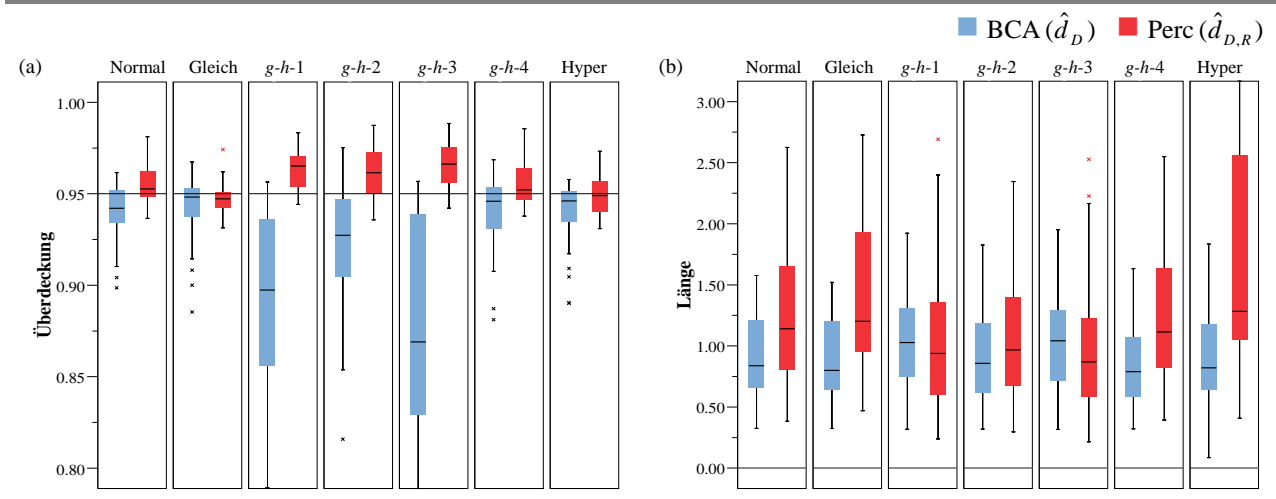


Abbildung 37. Simulierte Überdeckungswahrscheinlichkeiten und Intervalllängen von $BCA(\hat{d}_D)$ und $Perc(\hat{d}_{D,R})$ getrennt nach Verteilungen

Die Ergebnismuster stimmen insgesamt mit denjenigen bei unabhängigen Gruppen überein: Die Schätzung der klassischen Parameter gelingt – mit den dargestellten Einschränkungen – nur bei Verteilungen mit kurzen oder normalen Tails. Bereits bei einer Verteilung ($g-h-2$), die hinsichtlich Schiefe und Exzeß der Charakteristik der Exponentialverteilung entspricht, unterschreiten die hier ausgewählten Schätzer jedoch die nominale Überdeckungswahrscheinlichkeit zum Teil sehr deutlich. Eine robuste Schätzung bei Verteilungen mit extremen Tails ($g-h-1$, $g-h-3$) ist mit den hier simulierten Verfahren nicht möglich.

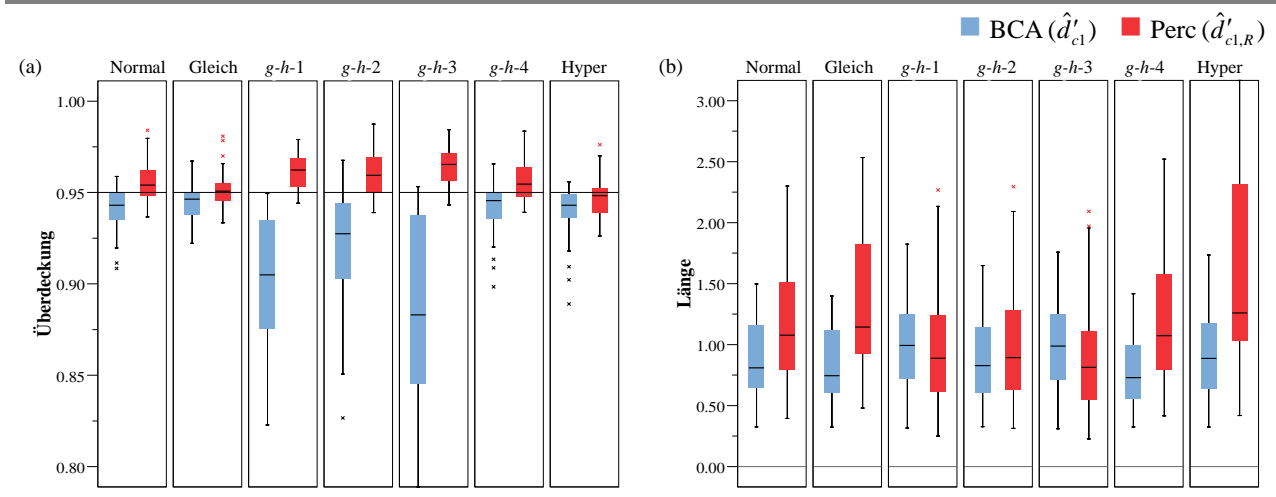


Abbildung 38. Simulierte Überdeckungswahrscheinlichkeiten und Intervalllängen von $BCA(\hat{d}'_{c1})$ und $Perc(\hat{d}'_{c1,R})$ getrennt nach Verteilungen

Im Gegensatz dazu gelingt die Schätzung der robustifizierten Parameter über alle hier realisierten Modelle hinweg in den meisten Fällen mit einer Überdeckungswahrscheinlichkeit nahe dem nominalen Niveau. Die Boxplots der Intervalllängen machen wiederum deutlich, daß die Schätzung robustifizierter Parameter bei Verteilungen mit kurzen oder normalen Tails nicht so präzise gelingt wie die Schätzung der klassischen Parameter.

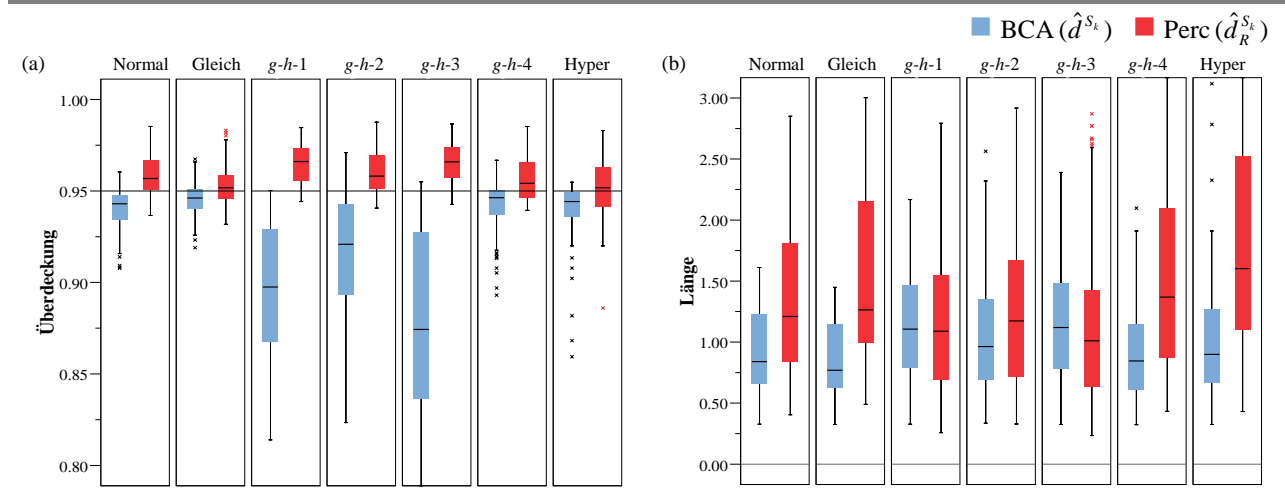


Abbildung 39. Simulierte Überdeckungswahrscheinlichkeiten und Intervallängen von $\text{BCA}(\hat{d}^{S_k})$ und $\text{Perc}(\hat{d}^{S_k})$ getrennt nach Verteilungen

2.2.5.3 Fazit

Die Ergebnisse der Simulationen zu standardisierten Lagedifferenzen bei verbundenen Stichproben zeigen analog zu den Ergebnissen bei unabhängigen Stichproben auf (vgl. Kapitel 2.1.5.3), daß die Einhaltung des Konfidenzniveaus der KQ-Schätzer vor allem von der zugrundeliegenden Verteilung sowie der Größe des zu schätzenden Parameterwertes abhängt. Die Homogenität der Varianzen spielt hier hingegen nur bei schiefen Verteilungen eine Rolle und die Höhe der Korrelation schließlich beeinflusst die Einhaltung des nominalen Niveaus kaum. Wie auch in Studie 1 wird insgesamt deutlich, daß die KQ-Parameter (d_D , d' und d^{S_k}) anhand der hier untersuchten Verfahren unter den betrachteten Verteilungen mit stärkeren Tails ($g-h-1$, $g-h-2$ und $g-h-3$) nicht robust geschätzt werden können. Die Schätzung der (20%-)robustifizierten Größen gelingt hingegen insbesondere mit dem Perzentil-Bootstrap über alle Verteilungen hinweg relativ robust. Zusammenfassend lassen sich hier folgende Empfehlungen machen:

- Sofern von Verteilungen ausgegangen werden kann, deren Tailstärken nicht oberhalb derjenigen der Normalverteilung liegen, empfiehlt sich die Schätzung der herkömmlichen Parameter (d_D , d' und d^{S_k}) anhand der nichtzentralen Intervallschätzer und bei größeren Stichproben auch anhand des BCA-Bootstraps. Für die Punktschätzungen bieten sich die biaskorrigierten Statistiken \hat{d}_D , \hat{d}'_{c1} und \hat{d}^{S_k} an.
- Sofern auch von Verteilungen mit stärkeren Tails auszugehen ist, verbietet sich die Schätzung der herkömmlichen Parameter. In diesem Fall sollte einer robustifizierten Schätzung der Parameter $d_{D,R}$, d'_R oder $d^{S_k}_R$ anhand des Perzentil-Bootstraps der Vorzug gegeben werden. Die Punktschätzung sollte anhand der t -korrigierten Statistiken $\hat{d}_{D,R}$, $\hat{d}'_{c1,R}$ und $\hat{d}^{S_k}_R$ erfolgen.

Wie im Falle unabhängiger Stichproben wird auch hier aufgrund des Effizienzverlustes der robustifizierten Schätzer unter Verteilungen mit kurzen Tails für diese keine generelle Empfehlung ausgesprochen. Hinsichtlich der weiteren Diskussion zu Konsequenzen für die praktische Verwendung der Verfahren sei auf das Fazit zu Studie 1 in Kapitel 2.1.5.3 sowie die Zusammenfassung der Arbeit in Kapitel 5 verwiesen.

3 Stochastische Überlegenheit

Als stochastische Überlegenheit wird im folgenden die Wahrscheinlichkeit bezeichnet, daß ein zufällig gezogener Wert aus einer Gruppe größer ist als ein zufällig gezogener Wert aus einer anderen Gruppe. Diese Wahrscheinlichkeit kann im Falle unabhängiger Gruppen anhand der Statistik U aus dem Test nach Mann und Whitney (1947) geschätzt werden über $U/(n_1 n_2)$. Als einer der ersten hat Birnbaum (1956) sich mit den Eigenschaften dieser Statistik beschäftigt und auch bereits Intervallschätzer vorgeschlagen. In der mathematischen Literatur wurde dieser generalisierten U -Statistik kein gesonderter Name gegeben und der zugehörige Wahrscheinlichkeitswert in der Regel als $\Pr(X < Y)$ notiert. Da hier die Gruppenbezeichnung über den Index vorgenommen wird und die Richtung des Gruppenunterschiedes analog zur Lagedifferenz definiert wird, sei die stochastische Überlegenheit für den bindungsfreien Fall mit $\Pr(X_1 > X_2)$ bezeichnet. In der psychologischen Literatur wird diese Statistik vergleichsweise spät (wieder)⁷ als Maß der Effektgröße diskutiert und zudem unter diversen Begriffen und Varianten verhandelt: Common Language Effect Size CL (McGraw & Wong, 1992), Dominance Statistic δ (Cliff, 1993), Probability of Superiority PS (Grissom, 1994), Measure of Stochastic Superiority A_{12} (Vargha & Delaney, 2000). Zudem ist $\Pr(X_1 > X_2)$ zu verschiedenen insbesondere in der medizinischen Literatur diskutierten Maßen äquivalent, so z.B. der Area Under the Receiver Operating Characteristic Curve $AUROC$ (Hanley & McNeil, 1982), dem Mean Ridit aus der Ridit Analyse (Bross, 1958) sowie Somers' D (Somers, 1962). Weitere eng verwandte Maße sind der Gini-Koeffizient, der c -Index nach Harrell sowie auch Agrestis α . Der Aufbau dieses Kapitels gleicht dem vorherigen zur standardisierten Lagedifferenz, d.h. es werden getrennt für unverbundene und verbundene Stichproben jeweils zunächst einige Eigenschaften der vorgeschlagenen Parameter berichtet, eine Auswahl geeigneter Schätzer vorgestellt und schließlich anhand von Simulationen untersucht.

3.1 Unverbundene Stichproben

Definitionen der stochastischen Überlegenheit unterscheiden sich vor allem darin, ob die angenommenen Modelle bindungsfrei sind oder nicht; und die Schätzungen dieser Parameter werden entweder parametrisch, semiparametrisch oder nichtparametrisch vorgenommen.

3.1.1 Varianten in der psychologischen Literatur

Mit einiger Verzögerung wurden die verschiedenen Ansätze zur Definition und Schätzung der stochastischen Überlegenheit auch in der Psychologie diskutiert, wobei sie als Effektgrößen zudem mit neuen Namen versehen wurden. Hier wird die Bezeichnung *stochastische Überlegenheit* aus der Arbeit von Vargha und Delaney (2000) übernommen, da diese Autoren den Parameter so definieren – nämlich als Wahrscheinlichkeitswert und unter Berücksichtigung von Bindungen –, wie er auch Gegenstand der vorliegenden Untersuchung ist. In der Mathematik hingegen spricht man im Deutschen eher von *relativem Effekt* (Brunner & Munzel, 2002) sowie von *stochastisch größer* und *kleiner*.

⁷ Da Wilcoxon (1945) einen äquivalenten Test entwickelt hatte – Unterschiede liegen darin, daß Wilcoxon nur auf den Fall gleich großer Gruppen eingeht und keine Stetigkeit voraussetzt –, wird der U -Test vielfach auch als Wilcoxon-Mann-Whitney-Test bezeichnet. Wie Kruskal (1957) aber berichtet, wurde bereits 1914 ein ebenfalls zum U -Test äquivalentes Verfahren von dem Psychologen Gustav Deuchler vorgestellt. Die von ihm entwickelte Statistik ist identisch mit Cliffs δ (Cliff, 1993).

3.1.1.1 Common Language Effect Size nach McGraw und Wong

Die von McGraw und Wong (1992) als Common Language Effect Size bezeichnete Statistik CL ist identisch mit der stochastischen Überlegenheit $\Pr(X_1 > X_2)$ im Falle normalverteilter Gruppen. Falls $X_1 \sim N(\mu_1, \sigma_1^2)$ und $X_2 \sim N(\mu_2, \sigma_2^2)$, gilt die Beziehung

$$CL = \Pr(X_1 > X_2) = \Phi\left(\frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right),$$

wobei Φ die kumulative Standardnormalverteilung ist (vgl. Downton, 1973; Owen, Craswell & Hanson, 1964). Für die Schätzung von CL schlagen McGraw und Wong die Ersetzung der Parameter in obiger Formel durch die entsprechenden Stichprobenkennwerte vor. Durch diese parametrische Definition der Effektgröße CL ergibt sich unmittelbar die Anknüpfung an die gebräuchliche Effektgröße d bzw. d' nach Cohen, denn für Normalverteilungen gilt:

$$CL = \Phi(d'/\sqrt{2}) \Leftrightarrow d' = \sqrt{2}\Phi^{-1}(CL).$$

Als wesentlichen Vorteil von CL gegenüber d' führen McGraw und Wong die größere Anschaulichkeit und leichtere Interpretierbarkeit dieses Wahrscheinlichkeitswertes an, der insbesondere auch von Nichtstatistikern verstanden werden könne. In Tabelle 49 wird die Beziehung zwischen d' und CL anhand einiger Werte unter der Normalverteilung beschrieben.

Tabelle 49. Beziehung zwischen d' und CL unter Normalverteilungen

d'	-2.0	-1.6	-1.2	-0.8	-0.4	0.0	0.4	0.8	1.2	1.6	2.0
CL	.079	.129	.198	.286	.389	.500	.611	.714	.802	.871	.921

McGraw und Wong führen also die stochastische Überlegenheit unter neuem Namen als Maß der Effektgröße in die psychologische Literatur ein. Eine mögliche Motivation für ihre Beschränkung auf das Normalverteilungsmodell mag sein, daß die Berechnung von CL einfach ist und insbesondere auch im nachhinein auf Basis von Maßzahlen erfolgen kann, die in nahezu allen Studien berichtet werden. Zudem ist die Darstellung der Beziehung zwischen stochastischer Überlegenheit und standardisierter Lagedifferenz unter der Normalverteilung für jeden Anwender sinnvoll, der mit Cohens Effektgrößen vertraut ist. Diese Beziehung erleichtert zum einen die Beurteilung eines CL -Wertes im jeweiligen Forschungsbereich, zum anderen aber macht diese auch deutlich, daß unter der Normalverteilung mit CL prinzipiell keine anderen Verteilungseigenschaften erfaßt werden als durch die etablierte Effektgröße nach Cohen.

Die schätztheoretischen Vorzüge der stochastischen Überlegenheit gegenüber d' und insbesondere d aber werden durch diese Beschränkung auf das Normalverteilungsmodell vollends aufgegeben. Denn während der Schätzer $U/(n_1 n_2)$ erwartungstreu und konsistent ist, zur Klasse der gleichmäßig besten erwartungstreuen Schätzer (uniformly minimal variance unbiased estimator) gehört (Birnbaum, 1956) und als Rangstatistik auch robust ist (Cliff, 1996a), zeigt CL weder für das Normalverteilungsmodell optimale Eigenschaften (Downton, 1973), noch kann CL unter irgendeiner anderen Verteilung die gesuchte Wahrscheinlichkeit unverzerrt schätzen. Denn anders als bei vielen Verteilungsparametern und auch Effektgrößen wie d , bei denen die Form der Stichproben-

verteilung wesentlich für das Schätzen und Testen ist, kommt es bei CL auf die Verteilung der Meßwerte bzw. der Grundgesamtheit an. Die für CL genutzte Beziehung zwischen standardisierter Lagedifferenz und stochastischer Überlegenheit gilt aber eben nur unter Normalverteilungen. McGraw und Wong zeigen zwar für einige Verteilungen, deren Schiefe zwischen -2 und $+2$ und deren Exzeß zwischen 0 und 14 liegen, daß die Abweichung zwischen dem gesuchten Parameter und dem mit \widehat{CL} geschätzten Wert gering ist; die Autoren geben hier einen maximalen Unterschied von $.06$ an. Doch bleibt diese Demonstration unbefriedigend, weil aus der Literatur zur Robustheit der kleinsten Quadratschätzer von vornherein bekannt ist (Wilcox, 2005), daß diese Abweichung durch die Wahl von anderen Verteilungen mit stärkeren Tails beliebig, d.h. bis zu dem maximalen Bias von $.5$, vergrößert werden kann. Sofern also nicht von Normalverteilung der Grundgesamtheit ausgegangen werden kann, bietet CL keine geeignete Möglichkeit zur Berechnung der stochastischen Überlegenheit.

3.1.1.2 Dominance Statistic nach Cliff

Der Ansatz von Cliff (1993, 1996a, 1996b) hingegen steht ganz in der Tradition der verteilungsfreien Statistik, die auch in der Psychologie vielfach als nichtparametrisch bezeichnet wird. Cliff geht davon aus, daß psychometrische Daten in aller Regel nicht das Niveau von Intervallskalen besitzen und daß zudem auch die Fragestellungen der psychologischen Forschung oft nur ordinalen Charakter haben. Insofern kommen ordinale Statistiken nicht nur als Alternative zu den klassischen Verfahren im Falle von Voraussetzungsverletzungen in Betracht, sondern stellen die zumeist adäquateren Verfahren zur Beantwortung der ohnehin nur ordinalen Fragestellungen dar. Als weiteren Vorteil der rangbasierten Statistiken sieht Cliff die höhere Robustheit und vielfach auch höhere Effizienz an. Vor diesem Hintergrund modifiziert Cliff nun die in der mathematischen Literatur für den stetigen Fall definierte stochastische Überlegenheit in zweierlei Hinsicht. Zum einen werden Rangbindungen berücksichtigt und zum anderen wird der Wertebereich auf -1 bis $+1$ aufgespannt. Cliff definiert die Dominanzstatistik, die im Original mit d bezeichnet wird, wie folgt:

$$\delta = \Pr(X_1 > X_2) - \Pr(X_1 < X_2).$$

Als Schätzer dient die Differenz zwischen dem Anteilswert, mit welchem Werte der ersten Gruppe größer sind als Werte der zweiten Gruppe und dem Anteilswert für den umgekehrten Vergleich:

$$\hat{\delta} = [\#(x_{1i} > x_{2j}) - \#(x_{1i} < x_{2j})] / (n_1 n_2),$$

wobei $\#$ die *Anzahl der Fälle* bezeichnet, in denen die Bedingung gegeben ist. Wenn alle Werte von X_2 größer sind als die Werte von X_1 , ist $\delta = -1$, im umgekehrten Fall $+1$ und im Falle der stochastischen Gleichheit ist $\delta = 0$. Für stetige, bindungsfreie Daten lautet die Beziehung zur stochastischen Überlegenheit $\Pr(X_1 > X_2) = (\delta + 1)/2$.

3.1.1.3 Probability of Superiority nach Grissom

Mit dem Hinweis darauf, daß die von McGraw und Wong eingeführte Statistik CL lange bekannt sei und lediglich einen neuen Namen bekommen habe, erinnert Grissom (1994) vor allem an die Möglichkeit, die stochastische Überlegenheit – die von ihm Probability of Superiority (PS) genannt wird – anhand der U -Statistik nichtparametrisch schätzen zu können. Ferner zeigt Grissom die

Beziehungen zwischen PS , der punkt-biseriellen Korrelation und verschiedenen standardisierten Lagedifferenzen auf. Schließlich empfiehlt Grissom die Berücksichtigung von PS in der primären sowie metaanalytischen Forschung. Einen eigenen Beitrag zur (Intervall-)Schätzung von PS macht Grissom jedoch nicht, weshalb diese Arbeit hier keine weitere Berücksichtigung findet.

3.1.1.4 Measure of Stochastic Superiority nach Vargha und Delaney

Eine weitere Variante der stochastischen Überlegenheit geht auf Vargha und Delaney (2000) zurück, wobei hier ebenfalls das Auftreten von Bindungen berücksichtigt wird und sich die resultierende Statistik von Cliffs δ nur durch die Normierung des Wertebereiches unterscheidet:

$$A_{12} = \Pr(X_1 > X_2) + \frac{1}{2}\Pr(X_1 = X_2).$$

Im stetigen Fall gilt also $A_{12} = \Pr(X_1 > X_2)$ und grundsätzlich gilt $A_{12} = (\delta + 1)/2$. Wie unten noch ausführlicher dargestellt wird, schlagen Vargha und Delaney eine parametrische Analyse der Rangstatistiken vor und geben für die Punktschätzung folgende Beziehung an:

$$\hat{A}_{12} = [\#(x_{1i} > x_{2j}) + \frac{1}{2}\#(x_{1i} = x_{2j})]/(n_1 n_2) = [\bar{R}_1 - (n_1 + 1)/2]/n_2,$$

wobei \bar{R}_1 der Mittelwert der Ränge aus Gruppe 1 ist und gebundene Ränge mit dem jeweiligen mittleren Rangwert versehen werden. Dieses Vorgehen entspricht dem für Bindungen modifizierten U -Test und bildet im folgenden die Grundlage für die Definition des Parameters und die Anpassung der Schätzer, die nur für den üblicherweise betrachteten bindungsfreien Fall entwickelt wurden.

3.1.2 Parameter und Punktschätzer

Da psychometrische Daten vielfach Bindungen aufweisen, wird im folgenden der als stochastische Überlegenheit bezeichnete Parameter $\theta = A_{12}$ betrachtet⁸. Ein erwartungstreuer Schätzer für θ ist die für Bindungen modifizierte Statistik nach Mann und Whitney (1947):

$$\hat{\theta} = \frac{U}{n_1 n_2} = \frac{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} U_{ij}}{n_1 n_2},$$

wobei $U_{ij} = 1$, falls $X_1 > X_2$, $U_{ij} = \frac{1}{2}$, falls $X_1 = X_2$ und $U_{ij} = 0$, falls $X_1 < X_2$. Die hierzu äquivalente Schätzung über die Rangstatistik ist im vorherigen Kapitel bereits angegeben und identisch mit dem von Brunner und Munzel (2002) als relativem Effekt bezeichneten Maß. Neben $\hat{\theta}$ gibt es eine Vielzahl für spezifische Verteilungsmodelle optimierte Schätzer der stochastischen Überlegenheit (vgl. Ali & Woo, 2005), die jedoch nicht Gegenstand der folgenden Untersuchungen sind; eine vergleichende Untersuchung der Eigenschaften von Punktschätzern wird für die stochastische Überlegenheit hier also nicht durchgeführt. Zur Erwartungstreue und Konsistenz von $\hat{\theta}$ siehe auch Brunner und Munzel (2002).

Die in Tabelle 49 dargestellte Beziehung zu d' weist bereits auf zwei Eigenschaften von θ hin, die generell gelten: Falls sich die beiden betrachteten Verteilungen (1) nicht oder (2) nur hinsicht-

⁸ Während der Buchstabe θ in Kapitel 2 der vorliegenden Arbeit einen beliebigen Parameter notiert hat, steht θ im folgenden für die stochastische Überlegenheit.

lich ihrer Varianz unterscheiden, ist $\theta = \frac{1}{2}$. Im Rahmen des Behrens-Fisher-Problems (vgl. Kapitel 1.5) ist ein Test der Hypothese $H_0 : \theta = \frac{1}{2}$ zugleich auch ein Test der Hypothese $H_0 : \mu_1 = \mu_2$. Der Test $H_0 : \theta = \frac{1}{2}$ gegen $H_1 : \theta \neq \frac{1}{2}$ wird daher auch als nichtparametrisches Behrens-Fisher-Problem bezeichnet (Brunner & Munzel, 2000; Neubert, 2006). Eine weitere wesentliche Eigenschaft ist schließlich (3), daß θ invariant ist unter allen ordnungserhaltenden Transformationen. Die stochastische Überlegenheit kann also sowohl bei metrischen als auch bei ordinalen Variablen jeweils mit und ohne Bindungen zum Einsatz kommen.

Die Robustheit der stochastischen Überlegenheit im Vergleich zu d kann an dem Beispiel der Abbildung 4 in Kapitel 2.1.1.6 (siehe S. 22) illustriert werden. Die Werte von θ lauten für Abb. (a) .71, (b) .68 und (c) .57. Vergleichbar mit der robustifizierten Lagedifferenz reagiert die stochastische Überlegenheit kaum auf die Extremwerte des kontaminierten Modells und führt somit auch zu dem Urteil, daß die Situationen (a) und (b) ähnlicher sind als (b) und (c). Die Werte der unmittelbar mit d verknüpften Effektgröße CL sind also für (a) .71, (b) .57 und (c) .57 und führen entsprechend zu demselben Urteil wie d ; schließlich machen diese deutlich, daß CL unter tailstarken Verteilungen nur wenig mit dem Wahrscheinlichkeitswert θ zu tun hat.

Als Nachteil der stochastischen Überlegenheit wird vielfach die Nichttransitivität angesehen (Grissom & Kim, 2001; Newcombe, 2006a), die bei multiplen Paarvergleichen zu paradoxen Ergebnisstrukturen führen kann, z.B. ergibt sich für die drei Gruppen $\mathbf{x}_1 = (1, 6, 8)'$, $\mathbf{x}_2 = (4, 5, 7)'$ und $\mathbf{x}_3 = (2, 3, 9)'$ folgendes: $\hat{\theta}_{12} = 5/9$, $\hat{\theta}_{23} = 6/9$, aber auch $\hat{\theta}_{31} = 5/9$. Eine mögliche Lösung dieses Dilemmas besteht darin, die Gruppen nicht paarweise miteinander zu vergleichen, sondern in Bezug zu einer anderen, gemeinsamen Referenzverteilung. Für die Konstruktion einer Referenzverteilung bietet sich, sofern keine Kontrollgruppe vorliegt, die Zusammenfassung aller gegebenen Werte an, also $\mathbf{x}_R = (1, 2, 3, 4, 5, 6, 7, 8, 9)'$. Im Vergleich zu dieser Verteilung ergeben sich folgende Schätzungen $\hat{\theta}_{1R} = 13.5/27$, $\hat{\theta}_{2R} = 14.5/27$ und $\hat{\theta}_{3R} = 12.5/27$, mittels derer eine Ordnung der Verteilungen beschrieben werden kann. Um auf Parameterebene Transitivität zu erreichen, wurden die klassischen Verfahren nur für Lokationsalternativen definiert (vgl. Kapitel 1.5). Transitivität läßt sich jedoch auch mit der allgemeineren Forderung erhalten, daß sich die Verteilungsfunktionen nicht kreuzen (Brunner & Munzel, 2002).

3.1.3 Intervallschätzer

Für die vorliegende Fragestellung zur Konstruktion von Konfidenzintervallen der stochastischen Überlegenheit wurde eine Vielzahl von Verfahren entwickelt (u.a. Birnbaum & McCarty, 1958; Govindarajulu, 1968; Halperin, Gilbert & Lachin, 1987; Lumelskii & Volkovich, 2007; Mee, 1990; Sen, 1967). Die Auswahl der hier berücksichtigten Verfahren orientiert sich an verschiedenen Kriterien. Zunächst sollten die in der psychologischen Literatur vorgestellten Verfahren berücksichtigt werden, da diese am ehesten Eingang in die psychologische Forschungspraxis erhalten. Hierzu gehören die bereits genannten Verfahren nach Vargha und Delaney sowie nach Cliff. Die Verfahren von Cliff werden neben dem Verfahren von Brunner und Munzel für die Testung der stochastischen Überlegenheit von verschiedenen Autoren (Neuhäuser, Lösch & Jöckel, 2007; Wilcox, 2005) als die robustesten insbesondere bei Heterogenität und gebundenen Daten angesehen.

Ferner orientiert sich die Auswahl an aktuellen Simulationsstudien (Feng & Cliff, 2004; Hess & Kromrey, 2004; Newcombe, 2006b; Zhou, 2008). In der wohl umfangreichsten Simulation zu Intervallschätzern der stochastischen Überlegenheit hat Newcombe (2006b) insgesamt acht Schätzer verglichen, von denen neben dem Verfahren von Mee (1990) zwei neue Vorschläge des Autors selbst (Methode 3 und Methode 5) sehr gute Eigenschaften aufweisen. Die Eignung der Methode 5

wird auch durch die Arbeit von Zhou (2008) bestätigt, der diese dem Verhalten des Bootstraps sowie der von ihm entwickelten Edgeworth-Reihe gegenüberstellt. Im Gegensatz zu den Untersuchungen der standardisierten Lagedifferenzen werden Bootstrap-Verfahren im folgenden nicht berücksichtigt. Wesentlicher Grund hierfür ist, daß die anderen hier ausgewählten Verfahren bereits bei sehr kleinen Stichproben recht gute Überdeckungsleistungen zeigen. Um zwischen diesen Verfahren nun entscheiden zu können, werden in der folgenden Studie Stichprobengrößen simuliert – die kleinste Zellbesetzung ist $n = 3$ –, für welche das Bootstrap nicht geeignet ist (Hess & Kromrey, 2004). Zudem liegt mit der von Zhou vorgeschlagenen Edgeworth-Reihe eine Approximation in geschlossener Form vor, die in seinen Simulationen bei mittleren und großen Stichproben ($60 \leq N \leq 200$) bessere Überdeckung als das Bootstrap selbst gezeigt hat. Schließlich wird mit dem Shift-Intervall ein eigener Vorschlag gemacht. Gemeinsam ist allen genannten Verfahren, daß sie auf der asymptotischen Normalverteilung der U -Statistik beruhen.

Vor der Beschreibung der ausgewählten Schätzer seien kurz die für Konfidenzintervalle von Wahrscheinlichkeitswerten gängigen Konstruktionsprinzipien beschrieben. Eine Interpretation des Konfidenzintervalls ergibt sich in Verbindung mit der Testtheorie: Ein Konfidenzintervall beinhaltet alle Parameter, deren Werte unter der Nullhypothese für den beobachteten Wert nicht zu ihrer Ablehnung führen (Rinne, 2003). Wichtig hierbei ist, daß die kritischen Werte der Tests eineindeutig von θ abhängen. Die Bezeichnungen der folgenden asymptotischen Schätzverfahren ergeben sich aus den jeweiligen Tests, aus deren Inversion sie resultieren. Die Umkehrung eines Wald-Tests der Form $(\hat{\theta} - \theta)/S_{\hat{\theta}}$ führt zu einem Wald-Intervall:

$$\hat{\theta} \pm c_{1-\alpha/2} S_{\hat{\theta}}.$$

Bei einem θ mit beschränktem Wertebereich führt dieses Intervall jedoch zu einem bei extremen Werten zu Grenzschätzungen, die außerhalb dieses Bereiches liegen können, zum anderen wird die Schiefe der Verteilung von $\hat{\theta}$ nicht berücksichtigt, der Punktschätzer liegt stets im Zentrum des Intervalls. Aus den Untersuchungen zur Schätzung der Erfolgswahrscheinlichkeit ist seit langem bekannt, daß Wald-Intervalle erst bei sehr großen Stichprobenumfängen und nicht zu extremen Werten zu brauchbaren Schätzungen führen. Eine deutliche Verbesserung dieses Intervalls ist mit der sogenannten Delta-Methode möglich. Hierbei wird der bereichsbeschränkte Schätzer aus dem Einheitsintervall $(0,1)$ z.B. über die Logit-Transformation $\ln[\hat{\theta}/(1-\hat{\theta})]$ auf die reelle Achse transformiert, auf welcher ein Wald-Intervall bestimmt wird:

$$\text{Logit}(\hat{\theta}) \pm c_{1-\alpha/2} \frac{S_{\hat{\theta}}}{\hat{\theta}(1-\hat{\theta})}.$$

Die Grenzen $[UG, OG]$ dieses Intervalls werden dann in die Skala von θ rücktransformiert über: $[\exp(UG)/(1 + \exp(UG)), \exp(OG)/(1 + \exp(OG))]$.

Die Inversion eines Score-Tests, dessen Statistik hier vereinfacht notiert wird als $(\hat{\theta} - \theta)/\sigma_{\hat{\theta}}$ zur Verdeutlichung, daß die Varianz nicht aus den beobachteten Daten geschätzt wird, sondern sich aus der Nullhypothese und dem Stichprobenumfang ergibt, führt zu einem Score-Intervall, welches erstmalig Wilson (1927) zur Schätzung eines Anteilswertes vorgeschlagen hat. Gesucht werden bei einem Score-Intervall also die beiden Parameter, für die folgende Gleichung gilt:

$$|\theta - \hat{\theta}| = c_{1-\alpha/2} \sigma_{\hat{\theta}}.$$

Sofern diese Gleichung – anders als bei Wilson – nicht in eine geschlossene Form gebracht werden kann, wird die Schätzung der Grenzen über Iterationen vorgenommen. Ähnlich wie die Logit-Wald-Intervalle führt auch das Score-Intervall zu deutlich besserem Überdeckungsverhalten als das Wald-Intervall. Bei beiden liegen die geschätzten Grenzen stets im Wertebereich von θ und die Schiefe der Verteilung drückt sich in asymmetrischen (nicht um die Punktschätzung zentrierten) Intervallen aus. Das Inversionsprinzip kann auch auf weitere Tests wie z.B. den Likelihood-Ratio-Test angewendet werden, ein Ansatz der hier jedoch nicht verfolgt wird. Schließlich kann ein Intervall auf Basis eines exakten Tests konstruiert werden, welches als Clopper-Pearson-Intervall bezeichnet wird und in Kapitel 3.2.2.1.1.5 für binomialverteilte Anteilswerte beschrieben wird.

3.1.3.1 Methode von Mee

Das Verfahren nach Mee stellt zum einen die Verknüpfung eines Jackknife-Varianzschätzers (Sen, 1967) mit der von Halperin, Gilbert und Lachin (1987) empfohlenen Intervallformel dar, zum anderen besteht der Beitrag von Mee in der Behandlung extremer Punktschätzungen. Tabelle 50 gibt die Formeln des Intervallschätzers von Mee an, Tabelle 51 verdeutlicht die Berechnung an einem Zahlenbeispiel. Im Original ist das Verfahren nur für stetige Variablen definiert; analog zu Newcombe (2006b) wird dieses hier mit der in Kapitel 3.1.2 eingeführten Funktion durchgeführt, bei welcher gebundenen Werten die .5 zugewiesen wird.

Der Unterschied zwischen der von Mee verwendeten Formel und der Anwendung eines Score-Intervalls kann wie folgt nachvollzogen werden. Die Varianz von $\hat{\theta}$ lautet nach Mee (1990, S. 794): $\theta(1-\theta)/N$, wobei hier $N = n_1 n_2 / [(n_1 - 1)p_1 + (n_2 - 1)p_2 + 1]$. Die Berechnung der Intervallenden resultiert aus der Umformung folgender Gleichung:

$$|\hat{\theta} - \theta| = z_{1-\alpha/2} \sqrt{\theta(1-\theta)/\hat{N}}.$$

Im Gegensatz zum Score-Intervall wird die Varianz aus den beobachteten Daten geschätzt, weshalb man dieses Intervall auch als Pseudo-Score-Intervall bezeichnet.

Tabelle 50. Intervallschätzer nach Mee (1990)

$$\hat{p}_1 = \sum_{i=1}^{n_1} \sum_{j=1, k \neq i}^{n_2} \frac{U_{ij} U_{kj}}{n_1 n_2 (n_1 - 1)}, \quad \hat{p}_2 = \sum_{i=1}^{n_1} \sum_{j=1, k \neq j}^{n_2} \frac{U_{ij} U_{ik}}{n_1 n_2 (n_2 - 1)};$$

$$b_1 = \frac{\hat{p}_1 - \hat{\theta}^2}{\hat{\theta} - \hat{\theta}^2}, \quad b_2 = \frac{\hat{p}_2 - \hat{\theta}^2}{\hat{\theta} - \hat{\theta}^2}; \quad A = \frac{(n_1 - 1)b_1 + 1}{1 - n_1^{-1}} + \frac{(n_2 - 1)b_2 + 1}{1 - n_1^{-1}}; \quad \hat{N} = \frac{n_1 n_2}{A};$$

$$C = \frac{z_{1-\alpha/2}^2}{\hat{N}}, \quad \text{wobei } z_{1-\alpha/2} \text{ das } 1-\alpha/2 \text{ Quantil der Standardnormalverteilung ist.}$$

$$D = \sqrt{C(\hat{\theta}(1-\hat{\theta}) + 0.25C)}.$$

Die Endpunkte des $1-\alpha$ Konfidenzintervalls für θ lauten: $\frac{\hat{\theta} + 0.5C \pm D}{1 + C}$.

Für den Fall $\hat{\theta} = 1$ oder $\hat{\theta} = 0$ sind \hat{N} , \hat{p}_1 und \hat{p}_2 nicht definiert. Zudem gelingt die Intervallschätzung nur sehr ungenau, wenn $\hat{\theta}$ nahe 1 oder 0 liegt (Mee, 1990, S. 799). Zur Stabilisierung der Schätzung schlägt Mee daher im Falle extremer Werte vor, die Berechnung von \hat{N} , \hat{p}_1 und \hat{p}_2 aufgrund schrittweise verschobener Daten vorzunehmen, für die $\min(\hat{\theta}, 1 - \hat{\theta}) \geq 0.5/\sqrt{n_1 n_2}$ erfüllt ist. Neben diesem ursprünglichen Vorschlag von Mee wird in den Simulationen zudem eine Variante von Newcombe (2006b) berücksichtigt, bei der nicht die Messwerte verschoben werden, sondern die Ränge, so daß die genannte Bedingung erfüllt ist. Der Vorteil hierbei ist, daß auch die Intervallschätzung invariant gegenüber (streng) monotoner Transformation ist, was bei Mees Ansatz nicht der Fall ist.

Tabelle 51. Intervallschätzung nach Mee (1990)

Matrix U_{ij}

		\mathbf{x}_2				
		28	32	38	47	51
\mathbf{x}_1	29	1	0	0	0	0
	30	1	0	0	0	0
	33	1	1	0	0	0

$$\hat{\theta} = \frac{4}{15} = .2\bar{6};$$

$$\hat{p}_1 = \frac{6}{30} = 0.2; \hat{p}_2 = \frac{2}{60} = 0.0\bar{3}; b_1 = 0.659; b_2 = -0.193;$$

$$A = 3.239; \hat{N} = 4.632; C = 0.829; D = 0.578.$$

Das 95%-Konfidenzintervall für θ lautet dann [.056; .688].

3.1.3.2 Cliffs Schätzer für die Dominanzstatistik

Gemeinsam ist den Verfahren von Cliff (1996a, 1996b) und Mee (1990) die Pseudo-Score-Formel zur Bestimmung der Intervallenden. Unterschiede zu Mee liegen zum einen in der Wahl des Varianzschätzers (vgl. Birnbaum, 1956), zum anderen in der Definition des zu schätzenden Parameters. Während Mee von stetigen Zufallsvariablen ausgeht, ist das Verfahren nach Cliff auch auf die Verrechnung von gebundenen Werten ausgerichtet. Die Formeln von Cliff werden hier im Original wiedergegeben (Tabelle 52), d.h. zur Schätzung eines Intervalls für δ . Für die Schätzung von θ lassen sich die Intervallgrenzen dann über $\theta = (\delta + 1)/2$ umrechnen. Der Rechenweg wird wiederum am Zahlenbeispiel in Tabelle 53 gezeigt. Zudem wird auch die Alternative von Delaney und Vargha (2002) berücksichtigt, insbesondere bei kleinen Stichproben einen kritischen t -Wert zu verwenden.

Tabelle 52. Intervallschätzer nach Cliff (1996a, 1996b)

$$D = \frac{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} D_{ij}}{n_1 n_2}, \text{ wobei } D_{ij} = \text{sign}(x_{1i} - x_{2j}), \text{ mit } i = 1, \dots, n_1 \text{ und } j = 1, \dots, n_2.$$

$$S_D^2 = \frac{(n_2 - 1)S_{D_i}^2 + (n_1 - 1)S_{D_j}^2 + S_D^2}{n_1 n_2}, \text{ wobei}$$

$$S_{D_i}^2 = \frac{\sum_{i=1}^{n_1} (D_i - D)^2}{n_1 - 1}, D_i = \frac{\sum_{j=1}^{n_2} D_{ij}}{n_2}, S_{D_j}^2 = \frac{\sum_{j=1}^{n_2} (D_j - D)^2}{n_2 - 1}, D_j = \frac{\sum_{i=1}^{n_1} D_{ij}}{n_1} \text{ und } S_{D_j}^2 = \frac{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (D_{ij} - D)^2}{(n_1 - 1)(n_2 - 1)}.$$

$$\text{Das Konfidenzintervall lautet: } \frac{D - D^3 \pm S_D z_{1-\alpha/2} \sqrt{(1 - D^2)^2 + S_D^2 z_{1-\alpha/2}^2}}{1 - D^2 + S_D^2 z_{1-\alpha/2}^2}.$$

$$\text{Alternativ mit } t_{\hat{v}, 1-\alpha/2} \text{ und } \hat{v} = \frac{(n_1 S_{D_i}^2 + n_2 S_{D_j}^2)^2}{(n_1 S_{D_i}^2)^2 + (n_2 S_{D_j}^2)^2}.$$

Da das Verfahren nach Cliff für $|D| = 1$ nicht definiert ist, wird dem Vorschlag von Neubert (2006) folgend als Varianzschätzung in diesem Fall derjenige Wert verwendet, der sich bei nahezu vollständig getrennten Verteilungen mit genau einem gebundenen Wert ergibt. Das von Feng und Cliff (2004) vorgeschlagene Verfahren wird hingegen nicht berücksichtigt, da das Intervall hierdurch zu stark geweitet wird und die Intervalllänge größer ist als bei ähnlichen Datensituationen, in denen die Varianz größer Null ist.

Tabelle 53. Intervallschätzung nach Cliff (1996)

Dominanzmatrix D_{ij}

		\mathbf{x}_2					
		28	32	38	47	51	D_i
\mathbf{x}_1	29	1	-1	-1	-1	-1	-0.6
	30	1	-1	-1	-1	-1	-0.6
	33	1	1	-1	-1	-1	-0.2
D_j		1	-0.3	-1	-1	-1	

$$D = \frac{-7}{15} = -.4\bar{6};$$

$$S_{D_i}^2 = \frac{0.10\bar{6}}{2} = 0.05\bar{3}; S_{D_j}^2 = \frac{3.0\bar{2}}{4} = 0.7\bar{5}; S_{D_j}^2 = \frac{11.7\bar{3}}{8} = 1.4\bar{6}; S_D^2 = \frac{4 \cdot 0.05\bar{3} + 2 \cdot 0.7\bar{5} + 1.4\bar{6}}{15} = 0.213.$$

Das 95%-Konfidenzintervall für δ auf Basis der z -Verteilung lautet somit $[-.904; .447]$. Für die stochastische Überlegenheit $\hat{\theta} = (-.4\bar{6} + 1)/2 = .2\bar{6}$ ist das gesuchte Intervall dann $[.048; .724]$.

3.1.3.3 Rang-Welch-Verfahren nach Vargha und Delaney

Aufbauend auf den Arbeiten von Zimmermann und Zumbo (1993) schlagen Vargha und Delaney (2000) für den Test sowie für die Schätzung von θ den Welch-Test vor, wobei die Varianzschätzung anhand der Ränge erfolgt. Als Intervallschätzer wählen sie das für Lagedifferenzen herkömmliche zentrale Intervall:

$$\hat{\theta} \pm t_{\hat{\nu}, 1-\alpha} S_{W_R},$$

$$\text{wobei } S_{W_R} = \frac{1}{n_1 + n_2} \sqrt{\frac{S_{R_1}^2}{n_1} + \frac{S_{R_2}^2}{n_2}} \text{ und } \hat{\nu} = \left(\frac{S_{R_1}^2}{n_1} + \frac{S_{R_2}^2}{n_2} \right) \Bigg/ \left(\frac{S_{R_1}^4}{n_1^2(n_1-1)} + \frac{S_{R_2}^4}{n_2^2(n_2-1)} \right).$$

Im Beispiel mit $\mathbf{x}_1 = (29, 30, 33)'$, $\mathbf{x}_2 = (28, 32, 38, 47, 51)'$ und den zugehörigen Rangwerten $\mathbf{R}_1 = (2, 3, 5)'$, $\mathbf{R}_2 = (1, 4, 6, 7, 8)'$ ergeben sich $S_{R_1}^2 = 2.333$, $S_{R_2}^2 = 7.700$ und die geschätzten Freiheitsgrade $\hat{\nu} = 6$. Der Schätzer des Standardfehlers ist dann $S_{W_R} = 0.190$ und führt mit $t_{6, .975} = 2.447$ und $\hat{\theta} = .2\bar{6}$ zu dem 95%-Konfidenzintervall $[-.199; .732]$ bzw. $[0; .732]$. Das Über- oder Unterschreiten des Wertebereiches von θ kann bei allen Intervallschätzern auftreten, die vom Typ eines Wald-Intervalls sind. Dies wird in der folgenden Studie jedoch nicht berücksichtigt, sondern wie im Beispiel demonstriert wird dann der jeweilige Randwert 0 oder 1 als Grenze gesetzt.

3.1.3.4 Test nach Brunner und Munzel

Im Modell des U -Tests wird unter der Nullhypothese die Identität zweier stetiger Verteilungen angenommen, welche gegen die Alternativhypothese getestet wird, daß sich die beiden Verteilungen nur hinsichtlich ihrer Lagen unterscheiden (vgl. Tabelle 1 in Kapitel 1.5). Insbesondere für den Vergleich varianzheterogener Verteilungen ist der U -Test jedoch nicht geeignet, sofern dieser nur auf Lageunterschiede bzw. auf Abweichungen von stochastischer Gleichheit reagieren soll. Dieses Ergebnis gilt sowohl für die als exakt bezeichnete Variante, bei welcher der Randomisationstest nach Fisher angewendet wird, als auch für die approximative Testung über die Standardnormalverteilung. Der von Vargha und Delaney (2000) verfolgte Weg, diesem Problem durch die Anwendung des Welch-Tests mit Rängen zu begegnen, ist jedoch aus theoretischer Perspektive abzulehnen (Cliff, 1993), da eine wesentliche Voraussetzung der Teststatistik verletzt ist, nämlich die Unabhängigkeit von Zähler und Nenner: Je weiter der beobachtete Wert von .5 entfernt liegt, desto kleiner ist die Streuung. Da sich dieser Ansatz jedoch empirisch teilweise bewährt hat (Zimmermann & Zumbo, 1993), wird er in der vorliegenden Untersuchung ebenfalls berücksichtigt. Brunner und Munzel (2000) konnten hingegen zeigen, daß unter zusätzlicher Verwendung der Ränge innerhalb der Gruppen eine asymptotisch normalverteilte Teststatistik für das nichtparametrische Behrens-Fisher-Problem entwickelt werden kann.

Das Verhalten des U -Tests sei anhand einer kleinen Simulation dem Rang-Welch-Test und dem Test nach Brunner und Munzel gegenübergestellt. Wie die empirischen α -Niveaus in Tabelle 54 verdeutlichen, reagiert der U -Test auf Heterogenität der Varianzen ganz ähnlich wie der Student- t -Test: Bei Gleichsinnigkeit von Heterogenität und Disbalance ist der U -Test konservativ, bei gleich großen Gruppen reagiert der Test relativ robust, bei gegenläufigem Verhältnis von Heterogenität und Disbalance ist jedoch auch der U -Test äußerst progressiv. Der Rang-Welch-Test (Vargha-Delaney) verhält sich zwar insgesamt robuster als der U -Test, neigt aber in allen Situationen zu

einem progressiven Testverhalten. Das Verfahren nach Brunner und Munzel hält hingegen in allen drei Situationen das nominale Niveau gut ein und erlaubt somit für diese Datensituationen eine robuste Testung der stochastischen Überlegenheit.

Tabelle 54. Empirische Fehlerwahrscheinlichkeiten $\hat{\alpha}^*$ unter Normalverteilungen mit $\theta = 1/2$, Varianzverhältnis 9 zu 1, $\alpha = .05$ und $N_{MC} = 50,000$

n_1, n_2	Mann-Whitney	Vargha-Delaney	Brunner-Munzel
20, 20	.067	.068	.051
30, 10	.017	.073	.052
10, 30	.125	.062	.054

Tabelle 55 gibt die Formeln zur Berechnung des Tests nach Brunner und Munzel (2000) sowie das in der Originalarbeit vorgeschlagene Intervall zur Schätzung von θ an. In ihrer Monographie zur nichtparametrischen Datenanalyse (Brunner & Munzel, 2002) empfehlen die Autoren hingegen die Verwendung des Logit-transformierten Wald-Intervalls. Wie bei dem Verfahren nach Cliff wird im Falle von $\hat{\theta} = 0$ oder $\hat{\theta} = 1$, wenn $S_{\hat{\theta}}^2$ nicht definiert ist, die minimal mögliche Varianz eingesetzt, die sich bei genau einer Bindung und ansonsten vollständig getrennten Gruppen für die gegebene Stichprobengröße ergäbe. Diese berechnet sich hier aus $S_{1,\min}^2 = (4n_1)^{-1}$ und $S_{2,\min}^2 = (4n_2)^{-1}$.

Tabelle 55. Intervallschätzer nach Brunner und Munzel (2000)

$$S_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} \left(R_{1i} - R_{1i}^{(1)} - \bar{R}_1 + \frac{n_1 + 1}{2} \right)^2; \quad S_2^2 = \frac{1}{n_2 - 1} \sum_{j=1}^{n_2} \left(R_{2j} - R_{2j}^{(2)} - \bar{R}_2 + \frac{n_2 + 1}{2} \right)^2; \quad S_{\hat{\theta}}^2 = \frac{S_1^2}{n_1 n_2^2} + \frac{S_2^2}{n_2 n_1^2};$$

$$t = \frac{\bar{R}_1 - \bar{R}_2}{(n_1 + n_2) S_{\hat{\theta}}} \quad \text{mit} \quad \hat{v} = \frac{(n_1 S_1^2 + n_2 S_2^2)^2}{\frac{(n_1 S_1^2)^2}{n_1 - 1} + \frac{(n_2 S_2^2)^2}{n_2 - 1}};$$

$$\hat{\theta} \pm t_{\hat{v}, 1-\alpha/2} S_{\hat{\theta}}.$$

Im Beispiel mit $\mathbf{x}_1 = (29, 30, 33)'$ und $\mathbf{x}_2 = (28, 32, 38, 47, 51)'$ lauten die Gesamt- und internen Ränge $\mathbf{R}_1 = (2, 3, 5)'$, $\mathbf{R}_2 = (1, 4, 6, 7, 8)'$ sowie $\mathbf{R}_1^{(1)} = (1, 2, 3)'$, $\mathbf{R}_2^{(2)} = (1, 2, 3, 4, 5)'$. Demnach ergeben sich als Rangmittelwerte $\bar{R}_1 = 3.3$, $\bar{R}_2 = 5.2$ und als Schätzer der stochastischen Überlegenheit $\hat{\theta} = .26$. Die Gruppenvarianzen sind $S_1^2 = 0.3$ und $S_2^2 = 1.7$, woraus der Standardfehler $S_{\hat{\theta}} = 0.205$ und die geschätzten Freiheitsgrade $\hat{v} = 4.862$ resultieren. Das Konfidenzintervall (Wald-Typ) mit $t_{4.862; .975} = 2.593$ heißt schließlich $[-.266; .799]$ bzw. $[0; .799]$.

Wie sich in diesem Datenbeispiel ohne Bindungen zeigt, stimmen die Varianzschätzungen nach Mee (1990) sowie nach Brunner und Munzel (2000) überein, d.h. es gilt: $\hat{N} = [\hat{\theta}(1 - \hat{\theta})] / S_{\hat{\theta}}^2$. Für den stetigen Fall ist der Test von Brunner und Munzel äquivalent zum Verfahren nach Mee. Sofern jedoch Bindungen auftreten, fällt die Varianzschätzung nach Mee größer aus als die Schätzung nach Brunner und Munzel. Eine vergleichende Untersuchung zur Güte der Varianzschätzer nach Mee, Brunner und Munzel sowie Cliff bei kleinen und mittleren Stichproben liegt bisher nicht vor und ist auch nicht Gegenstand der vorliegenden Studie.

3.1.3.5 Pseudo-Score-Intervall nach Newcombe

Da die Beziehung zwischen θ und $\sigma_{\hat{\theta}}^2$ kaum von der Verteilungsform der Daten abhängt, haben Hanley und McNeil (1982) vorgeschlagen, die Varianz unter der Annahme von Exponentialverteilungen zu berechnen, bei welchen $\hat{\theta}$ erschöpfend für die Schätzung der Varianz und also der Intervallgrenzen ist. Während diese Varianzschätzung zusammen mit einem Wald-Intervall in Newcombes Simulationen vergleichsweise schlecht abschneidet und hier nicht weiter verfolgt wird, führt folgende einfache Modifikation zu den empirisch besten Verfahren, die Newcombe in seiner Arbeit berücksichtigt hat. Hierbei wird die Varianz nicht anhand von $\hat{\theta}$ geschätzt, sondern folgende Gleichung jeweils für die obere und untere Intervallgrenze nach θ aufgelöst:

$$|\theta - \hat{\theta}| = z_{1-\alpha/2} \sigma_{\hat{\theta}}, \text{ wobei } \sigma_{\hat{\theta}}^2 = \frac{\theta(1-\theta)}{n_1 n_2} \left(1 + \frac{n_1^*(1-\theta)}{2-\theta} + \frac{n_2^* \theta}{1+\theta} \right).$$

Newcombe konstruiert also der Form nach ein Score-Intervall, welches wegen der erforderlichen Annahme der Exponentialverteilung ebenfalls als Pseudo-Score-Intervall bezeichnet wird. Eine geschlossene Formel, die sich analog zu Wilson (vgl. Kapitel 3.2.2.1.1.2) entwickeln ließe, wurde bisher nicht vorgestellt. Die Grenzen lassen sich über Intervallhalbierungsverfahren oder wie hier geschehen über sukzessive Approximation bestimmen. Newcombe führt zwei Varianten dieses Schätzers ein, wobei in Methode 3 $n_1^* = n_1 - 1$ und $n_2^* = n_2 - 1$, in Methode 5 $n_1^* = n_2^* = \frac{1}{2}(n_1 + n_2) - 1$ sind.

In dem genannten Zahlenbeispiel mit $\mathbf{x}_1 = (29, 30, 33)'$ und $\mathbf{x}_2 = (28, 32, 38, 47, 51)'$, in welchem $\hat{\theta} = .26$ ist, lauten die Lösungen nach Methode 3 für ein 95%-Konfidenzintervall [.072; .682]. Die zur unteren Grenze – genauer lautet der Wert $\theta_{UG} = .071692452723605$ – gehörige Streuung ist $\sigma_{\hat{\theta},UG} = .099478467707057$, so daß die Gleichung $\theta_{UG} + z_{.975} \sigma_{\hat{\theta},UG} = \hat{\theta}$ erfüllt ist. Entsprechendes gilt für die obere Grenze. Da die Gruppen verschieden groß sind, lautet das gesuchte 95%-Konfidenzintervall nach Methode 5 hier anders, und zwar [.065; .674].

3.1.3.6 Edgeworth-Reihe der U-Statistik

Aufgrund der Simulationen von Hess und Kromrey (2004), Zhou (2008) sowie einer eigenen Vorstudie wurde von der Aufnahme verschiedener Bootstrap-Verfahren für Konfidenzintervalle der stochastischen Überlegenheit Abstand genommen. Alternativ wird jedoch ein neuer Vorschlag von Zouh (2008) berücksichtigt, der eine Edgeworth-Reihe zur Approximation der Verteilung der U-Statistik entwickelt hat. Mit einer nach Edgeworth entwickelten Reihe läßt sich eine Verteilungsfunktion in Einheiten ihrer Kumulanten approximieren (Manoukian, 1986, S. 82). Voraussetzung hierfür ist, daß die betrachtete Statistik asymptotisch normalverteilt ist; Ziel hierbei ist es, diese Approximation zu beschleunigen. Das Verfahren nach Zhou für die studentisierte Mann-Whitney-U-Statistik sieht die in Tabelle 56 aufgeführten Berechnungen vor, wobei die Formeln für das Auftreten von Bindungen sowie die hier gewählte Richtung von θ angepaßt wurden. Im Beispiel mit $\mathbf{x}_1 = (29, 30, 33)'$ und $\mathbf{x}_2 = (28, 32, 38, 47, 51)'$ sind $\hat{g}_1 = (0.067, 0.067, -0.133)'$ und $\hat{g}_2 = (-0.733, -0.067, 0.267, 0.267, 0.267)'$, ferner sind $\hat{\gamma}_1 = -0.000593$, $\hat{\gamma}_2 = -0.0676$, $\hat{\gamma}_{12} = -0.000593$, so daß $\hat{h} = 0.449$, $\hat{\sigma} = 0.205$ und das Intervall [-.044; .762] bzw. [0; .762] lauten.

Tabelle 56. Edgeworth-Reihe nach Zhou (2008)

$$\hat{\sigma}^2 = \frac{1}{n_1(n_1-1)} \sum_{i=1}^{n_1} \hat{g}_1^2(x_{1i}) + \frac{1}{n_2(n_2-1)} \sum_{j=1}^{n_2} \hat{g}_2^2(x_{2j}),$$

$$\hat{h} = -\frac{z_{\alpha/2}^2}{2\hat{\sigma}^3} \left(\frac{\hat{\gamma}_1}{n_1^2} + \frac{\hat{\gamma}_2}{n_2^2} \right) + \frac{(z_{\alpha/2}^2 - 1)}{6\hat{\sigma}^3} \left(\frac{\hat{\gamma}_1}{n_1^2} + \frac{\hat{\gamma}_2}{n_2^2} + 6 \frac{\hat{\gamma}_{12}}{n_1 n_2} \right),$$

$$\hat{g}_1(x_{1i}) = n_2^{-1} \sum_{j=1}^{n_2} I(x_{1i} < x_{2j}) + 0.5I(x_{1i} = x_{2j}) - (1 - \hat{\theta}),$$

$$\hat{g}_2(x_{2j}) = n_1^{-1} \sum_{i=1}^{n_1} I(x_{1i} < x_{2j}) + 0.5I(x_{1i} = x_{2j}) - (1 - \hat{\theta}),$$

$$\hat{\gamma}_1 = n_1^{-1} \sum_{i=1}^{n_1} \hat{g}_1^3(x_{1i}), \quad \hat{\gamma}_2 = n_2^{-1} \sum_{j=1}^{n_2} \hat{g}_2^3(x_{2j}),$$

$$\hat{\gamma}_{12} = (n_1 n_2)^{-1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \hat{g}_1(x_{1i}) \hat{g}_2(x_{2j}) I(x_{1i} < x_{2j}) + 0.5I(x_{1i} = x_{2j}), \text{ wobei } I \text{ die Indikatorfunktion ist.}$$

Das zweiseitige Intervall lautet: $[\hat{\theta} - (z_{1-\alpha/2} - \hat{h})\hat{\sigma}, \hat{\theta} + (z_{1-\alpha/2} + \hat{h})\hat{\sigma}]$.

3.1.3.7 Shift-Konfidenzintervall

Angeregt durch die Datenverschiebung bei extremen Werten im Verfahren nach Mee und in Anlehnung an die nichtparametrische Schätzmethode von Bauer (1972) wurde schließlich ein eigenes, im folgenden Shift-Intervall bezeichnetes Verfahren in die Untersuchungen aufgenommen. Zunächst sei kurz die von Bauer für den Shift-Effekt ($\psi = \mu_1 - \mu_2$) zweier stetiger metrischer Variablen entwickelte Methode skizziert. Hierbei werden alle $n_1 n_2 = M$ vielen paarweisen Differenzen $\{\Delta_{ij} = x_{1i} - x_{2j}\}$ zwischen den beobachteten Meßwerten berechnet und aufsteigend sortiert: $\Delta_{(1)} \leq \Delta_{(k)} \leq \dots \leq \Delta_{(M)}$. Sodann werden die Werte einer beliebigen linearen Rangstatistik T an den Stellen $x_{1i} - \Delta, x_{2j}$ bestimmt, wobei Δ gerade links – durch diese weitere hinreichend kleine Verschiebung werden Bindungen ausgeschlossen – von der jeweiligen Ordnungsstatistik $\Delta_{(k)}$ liegt. So erhält man eine Folge von M Werten der Teststatistik, die auf den beobachteten Daten der einen Gruppe und den jeweils um Δ verschobenen Daten der anderen Gruppe basieren. Das gesuchte Konfidenzintervall des Shift-Effekts ist hier die Menge $\{\Delta | T(\Delta) \notin C_\alpha\}$, wobei C_α der Ablehnungsbereich des eingesetzten Tests ist. Die Konfidenzgrenzen für den Shift-Effekt ergeben sich also vereinfacht ausgedrückt aus den beiden extremsten Differenzen, um welche eine der beiden Gruppen verschoben werden kann und für die der Rangtest zur Beibehaltung der Nullhypothese führt. Da diese Methode nicht für eine feste Teststatistik formuliert wurde, kann diese mit geeigneter Wahl von T und C_α auch für den heteroskedastischen Fall eingesetzt werden (Neubert, 2006).

Um diesen Ansatz nun für die Konstruktion eines Intervalls der stochastischen Überlegenheit nutzbar zu machen, wurden zwei Modifikationen vorgenommen. Während bei Bauer nur die Δ -Werte von Interesse sind, für die sich Sprünge in der Funktion von T ergeben, sollen hier alle möglichen Verschiebungen vorgenommen werden, die zu verschiedenen Werten von $\tilde{\theta}$ – das sind die in den verschobenen Werten berechneten Schätzer für θ – führen. Im stetigen Fall ist hierfür genau eine weitere Verschiebung erforderlich, nämlich diejenige, die bei obiger Differenzbildung zu $\tilde{\theta} = 1$ führt. Ferner kann hier die Wahl der Konfidenzgrenzen nicht über einen Rangtest erfolgen,

sondern wird über Intervallschätzungen auf Basis der M verschobenen Datensituationen herbeigeführt. Als Grenzen des gesuchten Konfidenzintervalls werden die extremsten Werte von $\tilde{\theta}$ gewählt, deren Intervalle den beobachteten Wert möglichst exakt umschließen. Wie bei jeder diskreten Verteilung besteht jedoch auch hier das Problem, daß die Grenzen der anhand der verschobenen Gruppen berechneten Intervalle nicht genau auf dem beobachteten Wert $\hat{\theta}$ liegen, sondern in Abhängigkeit von der Stichprobengröße und Stetigkeit der zugrundeliegenden Verteilung mehr oder weniger weit davon entfernt sind. Aus diesem Grund werden drei Varianten vorgeschlagen, bei welchen die verschobenen Statistiken $\tilde{\theta}$ ausgewählt werden, deren Intervalle den beobachteten Wert (1) gerade noch überdecken, oder diejenigen Werte, deren Intervalle den beobachteten Wert (2) gerade nicht mehr umschließen und schließlich wird (3) eine Kontinuitätskorrektur dadurch versucht, daß ein gewichtetes Mittel der aus den ersten beiden Varianten hervorgegangenen Grenzen bestimmt wird. Die Gewichtung ergibt sich hier aus der Nähe der Intervallgrenzen zum beobachteten Wert.

Das genaue Vorgehen der drei Varianten soll am Datenbeispiel [$\mathbf{x}_1 = (29, 30, 33)'$ und $\mathbf{x}_2 = (28, 32, 38, 47, 51)'$] illustriert werden, bei welchem das Verfahren nach Brunner und Munzel (2000) zur Berechnung der verschobenen Intervalle herangezogen wird. Für die Berechnung eines Shift-Intervalls werden zunächst alle paarweisen M Differenzen $\Delta_{ij} = x_{1i} - x_{2j}$ der beobachteten Werte gebildet (vgl. Tabelle 57).

Tabelle 57. Indizes der beobachteten Werte ($i = \text{Gruppe 1}, j = \text{Gruppe 2}$), die paarweisen Differenzen Δ_{ij} , die Werte der stochastischen Überlegenheit $\tilde{\theta}$ sowie der zugeordneten (Wald-)Intervalle nach Brunner und Munzel (2000) für $\mathbf{x}_1 - \Delta_{\min} + \varepsilon$, \mathbf{x}_2 (Index 1) und $\mathbf{x}_1 - \Delta_{ij} - \varepsilon$, \mathbf{x}_2 (Indizes 2 bis 16)

Index	i	j	Δ_{ij}	$\tilde{\theta}$	$s_{\tilde{\theta}}^2$	\hat{v}	$t_{v,1-\alpha/2}$	UG	OG
1	1	5	-22	1.000	0.047	5.333	2.523	.878	1.122
2	1	5	-22	.933	0.094	5.333	2.523	.689	1.178
3	2	5	-21	.867	0.149	5.556	2.495	.480	1.253
4	1	4	-18	.733	0.205	4.862	2.593	.201	1.266
5	3	5	-18	.733	0.205	4.862	2.593	.201	1.266
6	2	4	-17	.667	0.221	4.745	2.613	.093	1.240
7	3	4	-14	.600	0.245	4.000	2.776	-.035	1.235
8	1	3	-9	.533	0.236	4.655	2.629	-.078	1.144
9	2	3	-8	.467	0.236	4.655	2.629	-.144	1.078
10	3	3	-5	.400	0.245	4.000	2.776	-.235	1.035
11	1	2	-3	.333	0.221	4.745	2.613	-.240	.907
12	2	2	-2	.267	0.205	4.862	2.593	-.266	.799
13	1	1	1	.133	0.149	5.556	2.495	-.253	.520
14	3	2	1	.133	0.149	5.556	2.495	-.253	.520
15	2	1	2	.067	0.094	5.333	2.523	-.178	.311
16	3	1	5	.000	0.047	5.333	2.523	-.122	.122

Mit Hilfe von Δ_{ij} lassen sich dann $M + 1$ viele Verschiebungen der Gruppen vornehmen, für welche alle Werte der stochastischen Überlegenheit, die innerhalb verschobener Gruppen und unter Vermeidung von Bindungen möglich sind, realisiert werden. Hierbei wird die Gruppe 1 für alle Differenzen nach $\mathbf{x}_1 - \Delta_{ij} - \varepsilon$ und für die minimale Differenz nach $\mathbf{x}_1 - \Delta_{\min} + \varepsilon$ verschoben und den beobachteten Werten der Gruppe 2 gegenübergestellt, wobei $0 < \varepsilon < |\Delta|_{\min}$ ist. Innerhalb dieser

$M + 1$ verschiedenen Konstellationen wird jeweils ein (Konfidenz-)Intervall für die stochastische Überlegenheit bestimmt. Tabelle 57 gibt die $M + 1$ vielen möglichen Konfidenzintervalle nach Brunner und Munzel (vgl. Tabelle 55) auf Basis gegeneinander verschobener Daten wieder.

Nach der beschriebenen Variante (1) sollen diejenigen Werte von $\tilde{\theta}$ als Grenzen festgelegt werden, deren zugehörigen Intervallgrenzen den hier beobachteten Wert von $\hat{\theta} = .26$ noch umschließen. Diese Bedingung wird von den verschobenen Stichproben mit den Indizes 4 und 15 erfüllt, somit lautet das erste Shift-Intervall $[.067; .733]$. Nach Variante (2) sollen die Intervallgrenzen aus den verschobenen Gruppen den beobachteten Wert gerade nicht mehr umschließen, so daß hiernach die Stichproben mit den Indizes 3 und 16 auszuwählen sind und zu dem Shift-Intervall $[0; .867]$ führen. Da diesen Strategien gemäß zu erwarten ist, daß die Variante (1) eher antikonservativ und Variante (2) eher konservativ überdecken, soll mit Variante (3) ein Kompromiß im Sinne einer Kontinuitätskorrektur herbeigeführt werden. Hierbei wird ein einfaches gewichtetes Mittel zwischen den nach Variante (1) und Variante (2) bestimmten Grenzen berechnet, wobei sich die Gewichte aus der Nähe der zugrundeliegenden Intervallgrenzen zum beobachteten Wert ableiten. Im Beispiel liegt die untere Grenze .201 aus der Stichprobe mit Index 4 näher am beobachteten Wert $\hat{\theta} = .26$ als die untere Grenze .480 aus der Stichprobe mit Index 3. Für die Mittelung der zugehörigen oberen Grenzen der Shift-Intervalle wird nun folgendes Gewicht bestimmt: $g = (UG_3 - \hat{\theta}) / (UG_3 - UG_4) = (.480 - .26) / (.480 - .201) = .775$. Die obere Grenze des gesuchten Shift-Intervalls nach Variante (3) lautet dann: $g \cdot OG_{v_1} + (1 - g) \cdot OG_{v_2} = .763$. Analog wird auch für die untere Grenze vorgegangen, so daß das gesuchte Intervall nach Variante (3) schließlich $[.053; .763]$ lautet.

Das beschriebene Vorgehen ist zwar in Anlehnung an Bauer konstruiert worden, hat aber bis auf die Verschiebetechnik nur noch wenig mit diesem gemein. Die Methode nach Bauer unterscheidet sich sowohl durch die Verwendung eines Rangtests als auch hinsichtlich des Ziels einer Schätzung für den Shift-Effekt (Neubert, 2006). Zu den asymptotischen Eigenschaften des hier vorgestellten Shift-Intervalls ist somit nichts bekannt. Als Voraussetzungen lassen sich aber die Stetigkeit der Daten sowie die Gültigkeit des Lokations-Skalen-Modells (vgl. Kapitel 2.1.4.3) nennen.

3.1.4 Notation der ausgewählten Statistiken

In der folgenden Studie wurden insgesamt 31 Intervallschätzer untersucht. Tabelle 58 gibt einen Überblick, wie die beschriebenen Ansätze zu dieser Vielzahl an Schätzern verknüpft wurden und wie sie notiert werden. Die Shift-Intervalle und der Schätzer MEE-PS-1 heben sich aufgrund der metrischen Verschiebetechnik von den anderen Verfahren ab, da sie nicht invariant gegenüber monotonen Transformationen sind und also für ordinale Variablen nur bedingt geeignet sind.

Tabelle 58. Intervallschätzer der Simulationsstudie zur stochastischen Überlegenheit bei unabhängigen Gruppen

Autor/Name	Formel/Verfahren	Notation
Mee (1990)	Pseudo-Score-Intervall mit Intervall-Shift	MEE-PS-1
	Pseudo-Score-Intervall mit Ordinal-Shift	MEE-PS-2
Cliff (1996b)	Pseudo-Score-Intervall mit z-Wert	CLI-PS-Z
	Pseudo-Score-Intervall mit t-Wert	CLI-PS-T
	Logit-Wald mit z-Wert	CLI-LOG-Z
	Logit-Wald mit t-Wert	CLI-LOG-T
Vargha & Delaney (2000)	Wald-Intervall mit z-Wert	VD-WALD
	Pseudo-Score-Intervall mit z-Wert	VD-PS-Z
	Pseudo-Score-Intervall mit t-Wert	VD-PS-T
	Logit-Wald mit z-Wert	VD-LOG-Z
	Logit-Wald mit t-Wert	VD-LOG-T
Brunner & Munzel (2000)	Wald-Intervall mit z-Wert	BM-WALD
	Pseudo-Score-Intervall mit z-Wert	BM-PS-Z
	Pseudo-Score-Intervall mit t-Wert	BM-PS-T
	Logit-Wald mit z-Wert	BM-LOG-Z
	Logit-Wald mit t-Wert	BM-LOG-T
Newcombe (2006b)	Pseudo-Score-Intervall mit $n_k^* = n_k - 1$	NEW-M-3
	Pseudo-Score-Intervall $n_k^* = \frac{1}{2}(n_1 + n_2) - 1$	NEW-M-5
Zhou (2008)	Edgeworth-Reihe	ZHOU
Shift-Intervall	Brunner-Munzel-z, Variante 1	S-BM-Z-1
	Brunner-Munzel-t, Variante 1	S-BM-T-1
	Brunner-Munzel-z, Variante 2	S-BM-Z-2
	Brunner-Munzel-t, Variante 2	S-BM-T-2
	Brunner-Munzel-z, Variante 3	S-BM-Z-3
	Brunner-Munzel-t, Variante 3	S-BM-T-3
	Cliff-z, Variante 1	S-CLI-Z-1
	Cliff-t, Variante 1	S-CLI-T-1
	Cliff-z, Variante 2	S-CLI-Z-2
	Cliff-t, Variante 2	S-CLI-T-2
	Cliff-z, Variante 3	S-CLI-Z-3
	Cliff-t, Variante 3	S-CLI-T-3

3.1.5 Verteilungsmodelle und Untersuchungsdesign

Hinsichtlich der verwendeten Verteilungen, der Erzeugung der Pseudozufallszahlen, der Datenmodelle, der Robustheitskriterien und der Anzahl der Simulationsdurchgänge ist die folgende Untersuchung identisch mit Studie 1 (vgl. Kapitel 2.1.4). Der Hauptunterschied zu den Untersuchungen der Lagedifferenzen liegt in der Variation des Stichprobenumfangs, denn die ausgewählten Verfahren versprechen bereits bei kleinen Stichproben akzeptable Schätzeigenschaften (Mee, 1990; Newcombe, 2006b).

Der Gesamtstichprobenumfang wurde mit acht Stufen zwischen $N = 12$ und $N = 40$ variiert, so daß für jede Stufe die Gruppengrößen in das Verhältnis 1:3 gesetzt werden konnten (12, 16, ..., 40). Wie bei den anderen Studien wurden die Effektgröße θ mit fünf Stufen (.500, .611, .714, .802, .871), das Varianzverhältnis $\sigma_1 : \sigma_2$ mit zwei Stufen (1:1 und 3:1) und das Verhältnis der Gruppengrößen $n_1 : n_2$ mit drei Stufen (1:1; 1:3 und 3:1) realisiert. Die genannten Variablen wurden nach

folgendem Schema gekreuzt, so daß in Studie 3 insgesamt 1400 Simulationsbedingungen umgesetzt wurden:

$$\begin{array}{c}
 \text{Verteilung} \\
 \left(\begin{array}{c}
 \text{Normal} \\
 \text{Gleich} \\
 g-h-1 \\
 g-h-2 \\
 g-h-3 \\
 g-h-4 \\
 \text{Hyper}
 \end{array} \right)
 \end{array}
 \times
 \begin{array}{c}
 N \\
 \left(\begin{array}{c}
 12 \\
 16 \\
 20 \\
 24 \\
 28 \\
 32 \\
 36 \\
 40
 \end{array} \right)
 \end{array}
 \times
 \begin{array}{c}
 \{\sigma_1 : \sigma_2; n_1 : n_2\} \\
 \left(\begin{array}{c}
 1:1; 1:1 \\
 1:1; 1:3 \\
 3:1; 1:1 \\
 3:1; 3:1 \\
 3:1; 1:3
 \end{array} \right)
 \end{array}
 \times
 \begin{array}{c}
 \theta \\
 \left(\begin{array}{c}
 .500 \\
 .611 \\
 .714 \\
 .802 \\
 .871
 \end{array} \right)
 \end{array}
 .$$

Die gewählten Parameterwerte von θ ergeben sich aus der Beziehung zu d' unter der Normalverteilung. Die in Studie 3 simulierten Normalverteilungsmodelle sind also identisch mit denjenigen aus Studie 1. Alle anderen Bedingungen unterscheiden sich hingegen hinsichtlich ihres Lageunterschiedes, da hier θ konstant gehalten wird. Um beispielsweise ein $\theta = .714$ – dieses korrespondiert unter Normalverteilungen mit $d' = 0.8$ – unter varianzhomogenen $g-h-3$ -Verteilungen zu erreichen, müssen diese um $.936$ gegeneinander verschoben werden; dieses Modell besitzt also die Effektgröße $d' = .936/1.687 = 0.555$ (siehe hierzu Anhang B).

3.1.6 Ergebnisse der Studie 3

Während das Verhalten der Intervallschätzer unter den verschiedenen stetigen Modellen sehr ähnlich ist, weichen die Überdeckungsraten unter dem negativen hypergeometrischen Modell von diesem zum Teil deutlich ab. Daher wird hier auf eine Gesamtanalyse verzichtet und jeweils eine Gegenüberstellung des (mittleren) Verhaltens unter den stetigen und des Verhaltens unter dem diskreten Modell vorgenommen. Auf die Unterschiede zwischen Verteilungen mit starken und schwachen Tails wird hier also nicht im einzelnen eingegangen, sondern auf die Tabellen im Anhang E verwiesen. Tabelle 59 gibt das Überdeckungsverhalten sowie die Intervalllängen aller simulierten Verfahren unter den sechs stetigen Modellen wieder, Tabelle 60 gibt die entsprechenden Werte für die negativen hypergeometrischen Verteilungen an. In Tabelle 59 werden also die Ergebnisse von insgesamt 1200 Versuchen zusammengefaßt, Tabelle 60 faßt die Ergebnisse der 200 diskreten, negativen hypergeometrischen Bedingungen zusammen.

Neben den bereits eingeführten Kennwerten zur Beurteilung der Überdeckung (vgl. Kapitel 2.1.5.2) werden für die Schätzung der Effektgrößen mit beschränkten Wertebereichen zudem die relativen Häufigkeiten angegeben, mit welchen der Parameter links von der unteren Intervallgrenze (LNCP) bzw. rechts von der oberen Intervallgrenze (RNCP) liegt (NCP steht für non-coverage probability). Da hier nur symmetrische Intervalle untersucht werden, sollte sich die simulierte Wahrscheinlichkeit des Überdeckungsfehlers $\hat{\alpha}^*$ gleichmäßig auf die beiden Seiten aufteilen. Als ein wesentliches Kriterium für die Beurteilung der Verfahrensgüte wird wiederum die mittlere absolute Abweichung des empirischen vom nominalen Überdeckungsniveau ($|\overline{CP} - .95|$) herangezogen. Unter den stetigen Verteilungen schneidet diesbezüglich das Shift-Verfahren S-BM-Z-1 mit $.009$ und unter den negativen hypergeometrischen Modellen MEE-PS-1 mit $.015$ am besten ab. Ein weiterer Vorteil von MEE-PS-1 ist die relativ gleichmäßige Aufteilung des Fehlerniveaus auf die beiden Intervallseiten; eine ähnlich gute Symmetrie erzielt nur CLI-LOG-Z.

Tabelle 59. Überdeckung und Intervalllänge unter stetigen Verteilungen

	Überdeckung								Länge	
	$pCP_{.936}^{.966}$	$pCP_{.936}^{1.000}$	$ CP - .95 $	Mittel	Min	Max	LNCP	RNCP	Mittel	Max
MEE-PS-1	.667	.799	.015	.946	.805	.986	.028	.026	0.406	0.670
MEE-PS-2	.653	.773	.017	.943	.792	.986	.031	.026	0.404	0.662
CLI-PS-Z	.660	.769	.021	.941	.678	.987	.035	.024	0.413	0.667
CLI-PS-T	.403	.846	.023	.955	.765	.997	.028	.017	0.445	0.769
CLI-LOG-Z	.382	.934	.021	.965	.829	.999	.017	.018	0.446	0.763
CLI-LOG-T	.133	.977	.029	.977	.881	1.000	.011	.012	0.487	0.871
VD-WALD	.482	.700	.022	.946	.711	1.000	.037	.017	0.469	0.861
VD-PS-Z	.531	.971	.015	.963	.917	1.000	.007	.030	0.432	0.698
VD-PS-T	.133	.996	.025	.975	.930	1.000	.004	.021	0.472	0.839
VD-LOG-Z	.248	.991	.023	.972	.923	1.000	.004	.024	0.479	0.881
VD-LOG-T	.062	<u>1.000</u>	.032	.982	.937	1.000	.002	.016	0.527	0.964
BM-WALD	.345	.345	.047	.903	.561	.955	.083	.014	0.470	0.875
BM-PS-Z	.614	.628	.028	.927	.662	.979	.047	.026	0.401	0.647
BM-PS-T	.549	.813	.023	.949	.680	.997	.034	.017	0.444	0.792
BM-LOG-Z	.599	.830	.020	.950	.760	.994	.029	.021	0.429	0.727
BM-LOG-T	.209	.928	.025	.968	.780	1.000	.019	.013	0.483	0.881
NEW-M-3	.502	.791	.020	.951	.840	.996	.015	.035	0.394	0.584
NEW-M-5	.475	.758	.022	.948	.818	.996	.016	.036	0.393	0.586
ZHOU	.401	.401	.049	.903	.558	.964	.080	.017	0.433	0.788
S-BM-Z-1	<u>.831</u>	.905	<u>.009</u>	.952	.901	.978	.012	.036	0.402	0.592
S-BM-T-1	.376	<u>1.000</u>	.019	.969	.938	.997	.007	.024	0.448	0.730
S-BM-Z-2	.604	.967	.014	.961	.917	.991	.010	.028	0.423	0.663
S-BM-T-2	.157	<u>1.000</u>	.025	.975	.942	.999	.006	.019	0.469	0.785
S-BM-Z-3	.750	.943	.011	.957	.911	.988	.011	.032	0.412	0.625
S-BM-T-3	.227	<u>1.000</u>	.023	.973	.939	.998	.006	.021	0.459	0.758
S-CLI-Z-1	.701	.943	.012	.958	.915	.986	.010	.032	0.413	0.614
S-CLI-T-1	.308	.998	.021	.970	.926	.999	.006	.023	0.446	0.688
S-CLI-Z-2	.454	.984	.018	.966	.921	.994	.009	.025	0.434	0.685
S-CLI-T-2	.161	.999	.026	.976	.933	.999	.005	.019	0.467	0.747
S-CLI-Z-3	.568	.971	.015	.963	.919	.992	.009	.028	0.423	0.648
S-CLI-T-3	.218	.999	.024	.973	.932	.999	.006	.021	0.456	0.714

Anmerkung: Unterstreichungen kennzeichnen die besten Resultate je Kriterium

Da S-BM-Z-1 unter den diskreten Modellen jedoch extrem schlecht überdeckt, wird alternativ das Shift-Verfahren S-BM-Z-3 im folgenden näher analysiert; dieses überdeckt im stetigen Fall ähnlich gut wie S-BM-Z-1 und zeigt auch im diskreten Fall ein relativ robustes Verhalten. Da die Vorteile der anderen Shift-Verfahren im Vergleich zu S-BM-Z-3 jeweils nur für ein Überdeckungskriterium spezifisch sind, werden auch diese im folgenden nicht näher dargestellt.

Hinsichtlich der Häufigkeit, mit der Überdeckungswahrscheinlichkeiten oberhalb von .936 zu beobachten sind ($pCP_{.936}^{1.000}$), schneidet der von Mee (1990) entwickelte Algorithmus (MEE-PS-1) tendentiell besser ab als die von Newcombe (2006b) zur Wahrung der Invarianz gegenüber monotonen Transformationen vorgeschlagene Variante (MEE-PS-2). Da die simulierte Überdeckungswahrscheinlichkeit von MEE-PS-1 auch unter den diskreten Verteilungen näher am nominalen Niveau liegt (vgl. $|CP - .95|$ in Tabelle 60), wird MEE-PS-2 in der graphischen Analyse nicht weiter berücksichtigt.

Tabelle 60. Überdeckung und Intervalllänge unter negativen hypergeometrischen Verteilungen

	Überdeckung								Länge	
	$pCP_{.936}^{.966}$	$pCP_{.936}^{1.000}$	$ \overline{CP} - .95 $	Mittel	Min	Max	LNCP	RNCP	Mittel	Max
MEE-PS-1	<u>.630</u>	.895	<u>.015</u>	.953	.769	.983	.022	.025	0.412	0.681
MEE-PS-2	.615	.845	.017	.949	.764	.983	.026	.025	0.409	0.676
CLI-PS-Z	.620	.730	.024	.936	.744	.983	.038	.025	0.406	0.655
CLI-PS-T	.450	.830	.025	.952	.764	.996	.030	.018	0.442	0.776
CLI-LOG-Z	.490	.905	.021	.960	.776	.997	.020	.020	0.437	0.761
CLI-LOG-T	.150	.940	.028	.973	.888	1.000	.013	.013	0.482	0.880
VD-WALD	.430	.610	.028	.935	.762	.998	.044	.021	0.458	0.854
VD-PS-Z	.415	.880	.021	.957	.834	.999	.009	.034	0.419	0.710
VD-PS-T	.120	.925	.028	.970	.850	1.000	.006	.023	0.460	0.848
VD-LOG-Z	.160	.915	.026	.966	.839	1.000	.007	.028	0.466	0.866
VD-LOG-T	.055	.935	.033	.977	.857	1.000	.004	.018	0.514	0.984
BM-WALD	.320	.320	.047	.904	.627	.966	.081	.015	0.469	0.880
BM-PS-Z	.580	.615	.030	.926	.720	.983	.047	.027	0.397	0.644
BM-PS-T	.510	.785	.025	.947	.765	.996	.035	.018	0.442	0.772
BM-LOG-Z	.570	.805	.022	.947	.758	.996	.032	.022	0.424	0.722
BM-LOG-T	.250	.910	.026	.966	.783	.997	.020	.014	0.479	0.863
NEW-M-3	.470	.810	.025	.950	.763	.996	.016	.034	0.394	0.584
NEW-M-5	.455	.770	.028	.946	.763	.996	.018	.036	0.393	0.585
ZHOU	.400	.440	.046	.908	.598	.983	.075	.016	0.433	0.812
S-BM-Z-1	.000	.000	.137	.813	.601	.903	.052	.135	0.312	0.481
S-BM-T-1	.030	.035	.097	.853	.719	.968	.035	.112	0.360	0.656
S-BM-Z-2	.055	<u>.990</u>	.037	.986	.859	1.000	.004	.010	0.520	0.862
S-BM-T-2	.030	<u>.990</u>	.041	.991	.920	1.000	.002	.007	0.565	0.983
S-BM-Z-3	.535	.790	.020	.953	.776	.996	.012	.035	0.413	0.620
S-BM-T-3	.290	.930	.024	.966	.793	1.000	.008	.026	0.460	0.759
S-CLI-Z-1	.000	.000	.123	.827	.658	.922	.047	.126	0.323	0.514
S-CLI-T-1	.060	.060	.097	.853	.710	.958	.036	.111	0.355	0.604
S-CLI-Z-2	.040	<u>.990</u>	.039	.988	.913	1.000	.003	.009	0.530	0.893
S-CLI-T-2	.030	<u>.990</u>	.041	.991	.929	1.000	.002	.007	0.560	0.946
S-CLI-Z-3	.445	.845	.022	.958	.779	.997	.011	.032	0.423	0.649
S-CLI-T-3	.230	.920	.027	.967	.792	.999	.008	.025	0.455	0.707

Anmerkung: Unterstreichungen kennzeichnen die besten Resultate je Kriterium

Neben diesen beiden Verfahren (S-BM-Z-3 und MEE-PS-1) werden ferner NEW-M-3 und VD-PS-Z näher betrachtet. NEW-M-3 ist im Hinblick auf die Nähe zum nominalen Niveau ($|\overline{CP} - .95|$) zwar verschiedenen anderen Verfahren unterlegen, es hat aber den wesentlichen Vorteil, daß es nur die Kenntnis der Punktschätzung und der Stichprobengröße erfordert. Entgegen dem Fazit von Newcombe, der NEW-M-5 als generell empfehlenswertes Verfahren beurteilt, schneidet hier NEW-M-3 sowohl im Hinblick auf den Abstand von der nominalen Wahrscheinlichkeit ($|\overline{CP} - .95|$) als auch hinsichtlich der Anzahl der Intervalle mit Überdeckungswahrscheinlichkeit $\geq .936$ in der Regel besser ab und wird daher für die weitere Analyse vorgezogen. Zu beachten ist, daß diese beiden Verfahren zu identischen Intervallen führen, wenn die Gruppen gleich groß sind. Wie Tabelle E-7 im Anhang zeigt, überdeckt NEW-M-5 im homoskedastischen Fall besser als NEW-M-3. Da dieser Vorteil jedoch relativ gering ist, wird hier NEW-M-3 insgesamt als die empfehlenswerte Methode erachtet, da sie im Falle heterogener Varianzen deutlich besser überdeckt. Mit Ausnahme der Verfahren nach Mee und nach Newcombe gilt für alle anderen, daß die Überdeckungsraten

unter den diskreten Modellen niedriger ausfallen als unter den stetigen. Zu den hier nicht im Detail dargestellten Verfahren ist ferner folgendes anzumerken. Das Pseudo-Score-Intervall und insbesondere die Logit-Transformation führen mit den Varianzschätzern nach Vargha und Delaney sowie nach Brunner und Munzel (2000) im Vergleich zur Wald-Methode zu konservativeren Schätzungen; die Anteile von Überdeckungswahrscheinlichkeiten $\geq .936$ nehmen jeweils deutlich zu. Sämtliche Verfahren mit der Varianzschätzung nach Brunner und Munzel zeigen unbalancierte Überdeckungsfehler, LNCP ist jeweils größer als RNCP; das bedeutet, daß die Intervalle tendentiell zu weit außen – hier also oberhalb des Parameters – liegen. Diese Disbalance zeigt sich auch für VD-WALD, bei den anderen VD-Verfahren hingegen verkehrt sich dieses Verhältnis. Nur die Varianzschätzung nach Cliff (1996a, 1996b) zeigt in Verbindung mit der Logit-Transformation sehr gut zentrierte Intervalle, LNCP und RNCP sind ungefähr gleich. Die Verfahren nach Cliff führen im Vergleich zu den Intervallen mit der Varianz nach Brunner und Munzel durchweg zu höheren Überdeckungsraten. Während die Varianzschätzung nach Brunner und Munzel im Vergleich zur Verwendung des z -Wertes in Verbindung mit kritischen t -Werten zu einer besseren Überdeckung hinsichtlich der absoluten Abweichung vom nominalen Niveau ($|\overline{CP} - .95|$) führt, ist dieses bei der Varianzschätzung nach Cliff umgekehrt. Für alle ausgewählten Varianzschätzer schließlich gilt, daß die Logit-Transformation zu durchschnittlich längeren Intervallen führt als das Pseudo-Score-Intervall. Das Verfahren nach Zhou (2008) zeigt unter stetigen Verteilungen mit durchschnittlich 5% Abweichung vom nominalen Niveau das insgesamt schlechteste Überdeckungsverhalten, unter den diskreten Verteilungen schneiden diesbezüglich nur die nach Variante (1) konstruierten Shift-Intervalle noch schlechter ab. Da das Verfahren nach Zhou für den hier untersuchten Stichprobenraum hinsichtlich keiner der berücksichtigten Kriterien Vorteile bietet, wird dieses hier nicht weiter berücksichtigt.

Bevor nun die vier ausgewählten Verfahren weiter gegenübergestellt werden, sei kurz auf die Bedeutung der Varianzschätzung bei extremen Werten eingegangen, wie sie bereits von Mee (1990) hervorgehoben wurde. Da die Varianzschätzer bei stetigen Verteilungen sowie die Pseudo-Score-Intervallformel identisch sind, gehen die Unterschiede zwischen MEE-PS-1 und BM-PS-Z allein auf den Effekt der (Shift-)Korrektur bei extremen Werten nahe Null oder Eins im Verfahren nach Mee zurück. Dieser Effekt sei anhand von Abbildung 40 illustriert, welche sämtliche Versuche unter den stetigen Verteilungen zusammenfaßt.

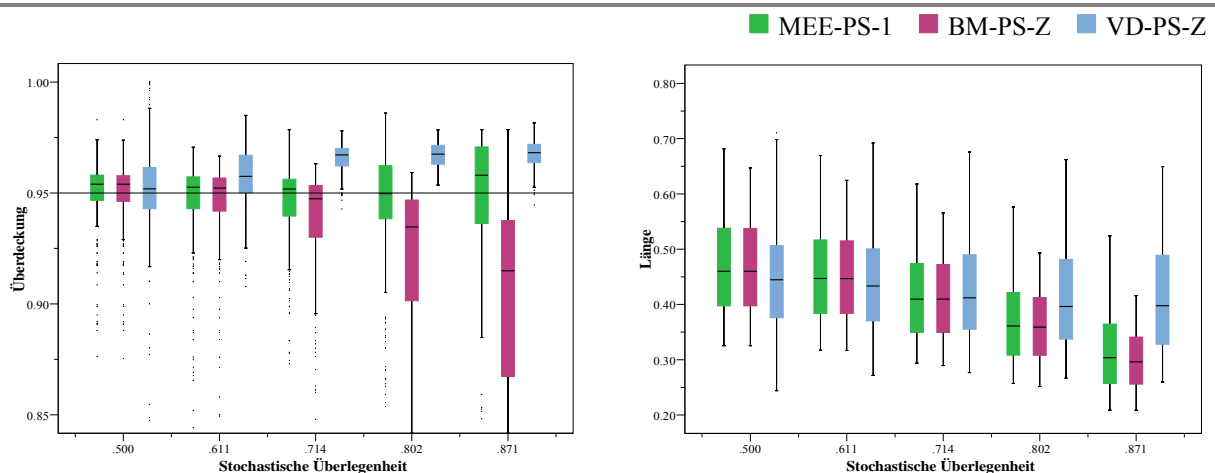


Abbildung 40. Simulierte Überdeckungswahrscheinlichkeiten und Intervalllängen von MEE-PS-1, BM-PS-Z und VD-PS-Z unter stetigen Verteilungen

Während die mediane Überdeckung von MEE-PS-1 über alle Simulationsbedingungen hinweg auch bei größeren Parameterwerten auf oder oberhalb des nominalen Konfidenzniveaus liegt, nimmt die Überdeckungswahrscheinlichkeit von BM-PS-Z mit steigendem Parameterwert kontinuierlich ab. Obwohl der Varianzschätzer nach Mee bzw. Brunner und Munzel (2000) deutlich bessere Schätzeigenschaften hat als der Schätzer des Verfahrens nach Vargha und Delaney (2000), ist die Überdeckungsleistung von VD-PS-Z besser. Die Ursache hierfür ist der zunehmend positive Bias, den der VD-Varianzschätzer mit steigenden Parameterwerten hat. Dieser führt zu relativ langen Intervallen bei extremeren Werten.

Das Verhalten der vier ausgewählten Verfahren sei zunächst in Tabelle 61 anhand zweier Kriterien, der mittleren absoluten Abweichung und der mittleren Intervalllänge, getrennt nach der Stichprobengröße und dem Parameterwert gegenübergestellt. Je Zeile sind die günstigsten Werte beider Kriterien unterstrichen. Unter den stetigen Verteilungen besitzt demnach das Shift-Intervall S-BM-Z-3 insbesondere bei kleinen Stichproben und auch bei großen Parameterwerten das beste Überdeckungsverhalten hinsichtlich $|\overline{CP} - .95|$. Unter der negativen hypergeometrischen Verteilung fällt die Entfernung zum nominalen Niveau beim Shift-Intervall jedoch durchweg größer aus. Hier zeigt sich insbesondere das Verfahren nach Mee als relativ unempfindlich gegenüber dem Auftreten von Bindungen. Während bei den Verfahren MEE-PS-1 und S-BM-Z-3 in aller Regel die Überdeckung bei größerem Umfang besser gelingt, ist dies bei den anderen beiden Verfahren und insbesondere bei den größeren Parameterwerten nicht der Fall. Aussagen zur Asymptotik der Verfahren lassen jedoch die hier betrachteten Stichprobengrößen bis $N = 40$ nicht zu.

Tabelle 61. Mittlere absolute Abweichung der simulierten von der nominalen Überdeckungswahrscheinlichkeit und Intervalllänge ausgewählter Verfahren

	N	θ	$ \overline{CP} - .95 $				Mittlere Länge			
			MEE-PS-1	VD-PS-Z	NEW-M-3	S-BM-Z-3	MEE-PS-1	VD-PS-Z	NEW-M-3	S-BM-Z-3
Stetig	12 bis 24	0.500	0.020	0.016	0.027	<u>0.014</u>	0.529	0.512	<u>0.489</u>	0.512
		0.611	0.019	0.017	0.023	<u>0.015</u>	0.513	0.505	<u>0.482</u>	0.505
		0.714	0.016	0.016	0.018	<u>0.015</u>	0.477	0.493	<u>0.458</u>	0.482
		0.802	0.019	0.015	0.015	<u>0.013</u>	0.425	0.484	<u>0.424</u>	0.448
		0.871	0.020	0.015	0.014	<u>0.011</u>	<u>0.368</u>	0.492	0.385	0.410
	28 bis 40	0.500	<u>0.004</u>	0.009	0.026	0.008	0.409	0.392	<u>0.382</u>	0.401
		0.611	<u>0.006</u>	0.007	0.023	0.009	0.395	0.384	<u>0.375</u>	0.391
		0.714	0.010	0.017	0.019	<u>0.009</u>	0.362	0.366	<u>0.352</u>	0.365
		0.802	0.016	0.018	0.017	<u>0.008</u>	<u>0.316</u>	0.347	0.317	0.327
		0.871	0.024	0.020	0.017	<u>0.008</u>	<u>0.263</u>	0.340	0.277	0.284
Hyper	12 bis 24	0.500	<u>0.025</u>	0.040	0.048	0.037	0.542	0.500	<u>0.487</u>	0.502
		0.611	0.020	<u>0.017</u>	0.029	0.025	0.522	0.498	<u>0.481</u>	0.497
		0.714	0.016	<u>0.014</u>	0.020	0.020	0.482	0.485	<u>0.458</u>	0.477
		0.802	0.018	0.016	<u>0.015</u>	0.017	0.429	0.469	<u>0.425</u>	0.450
		0.871	0.024	0.017	0.015	<u>0.014</u>	<u>0.380</u>	0.459	0.386	0.426
	28 bis 40	0.500	<u>0.009</u>	0.034	0.040	0.027	0.424	0.384	<u>0.380</u>	0.394
		0.611	<u>0.006</u>	0.016	0.032	0.017	0.402	0.380	<u>0.374</u>	0.387
		0.714	<u>0.009</u>	0.015	0.021	0.013	0.364	0.363	<u>0.352</u>	0.364
		0.802	<u>0.011</u>	0.020	0.014	0.013	<u>0.313</u>	0.338	0.318	0.330
		0.871	<u>0.013</u>	0.021	0.016	<u>0.013</u>	<u>0.260</u>	0.317	0.278	0.299

Anmerkung: Unterstreichungen kennzeichnen die besten Resultate je Kriterium

NEW-M-3 zeichnet sich vor allem durch die Kürze seiner Intervalle aus. Da Bindungen bei diesem Verfahren nicht berücksichtigt werden, treten hier die geringsten Unterschiede der Intervalllängen zwischen den stetigen und den diskreten Bedingungen auf. Während die Intervalle von VD-PS-Z unter dem diskreten Modell durchgängig etwas kürzer ausfallen, gibt es bei MEE-PS-1 und S-BM-Z-3 jeweils ein weniger einheitliches Bild. Die durchschnittlich längeren Intervalle bei größeren Parameterwerten können aber wiederum durch das Auftreten von Bindungen erklärt werden, durch welche die Verschiebung der Daten in beiden Verfahren in größeren Schritten erfolgt.

Abbildung 41 und Abbildung 42 zeigen nun getrennt für stetige und diskrete Verteilungen die simulierten Überdeckungswahrscheinlichkeiten der vier ausgewählten Verfahren. Vergleichbar mit den Ergebnissen zu Studie 1 fällt auch hier auf den ersten Blick auf, daß die Verfahren insgesamt bei Gegenläufigkeit von Disbalance und Heterogenität das nominale Konfidenzniveau am schlechtesten einhalten.

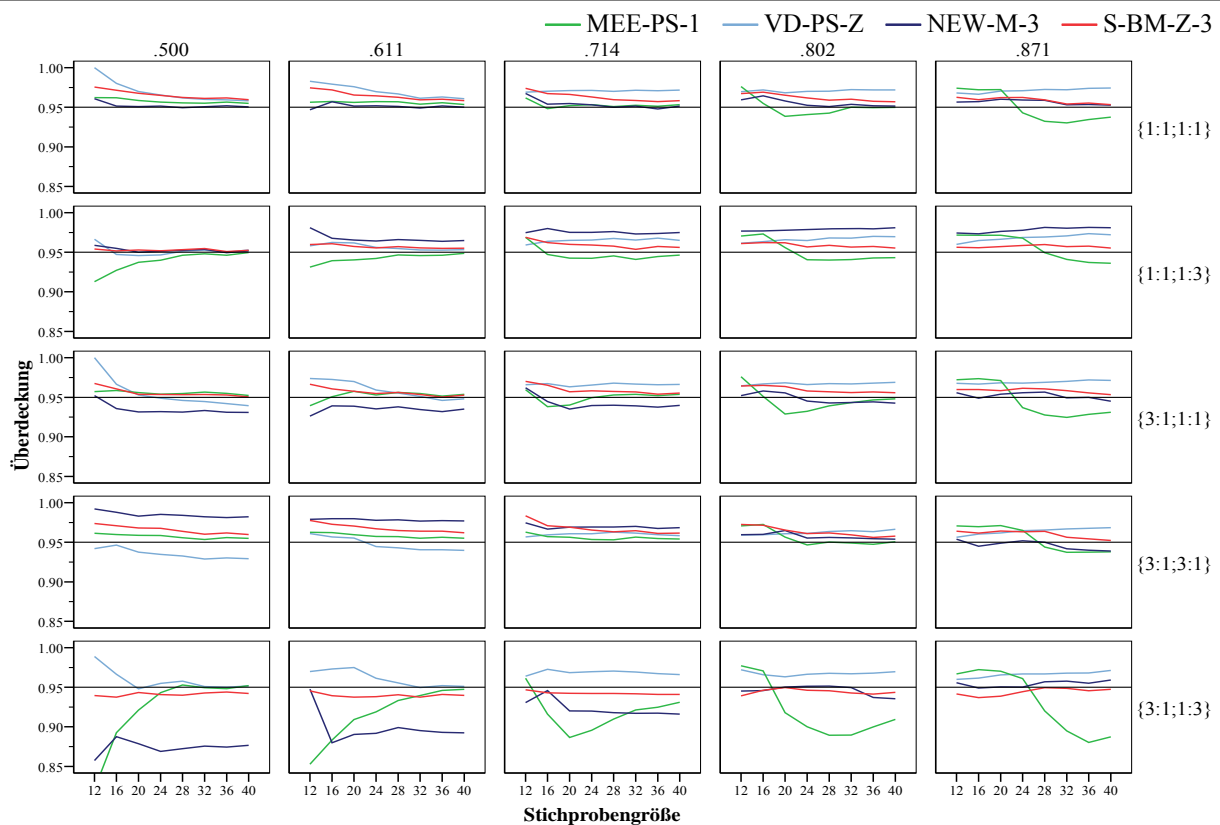


Abbildung 41. Mittlere simulierte Überdeckungswahrscheinlichkeiten ausgewählter Verfahren unter den stetigen Verteilungen

Mit Ausnahme der erwähnten Gegenläufigkeit von Gruppengröße und Varianz hält das Verfahren nach Mee unter den stetigen Verteilungen (vgl. Abbildung 41) bis zu einem Parameterwert von .714 die nominale Überdeckungswahrscheinlichkeit in aller Regel ein. Bei größeren Parameterwerten jedoch gerät die Überdeckung mit steigender Stichprobengröße aus dem konservativen in den progressiven Bereich. Während sich unter $\theta = .802$ die Überdeckungswahrscheinlichkeit bis $N = 40$ zumeist an das nominale Niveau wieder annähert, tritt dieser Effekt bei $\theta = .871$ später ein und ist unter allen Bedingungen auch bei $N = 40$ noch festzustellen, d.h. hier unterschreitet MEE-PS-1

unter allen Bedingungen das nominale Niveau. VD-PS-Z überdeckt unter den stetigen Verteilungen von allen ausgewählten Verfahren am konservativsten, in nur wenigen Situationen treten Unterschreitungen des nominalen Niveaus auf. Die größte Abweichung findet sich bei $\{.500; 3:1; 3:1\}$, hier lautet die empirische Überdeckung unter der Gleichverteilung $.917$ (vgl. Tabelle 59 und Tabelle E-3 im Anhang).

Die insgesamt größten Abweichungen vom nominalen Niveau treten auch beim Verfahren nach Newcombe (NEW-M-3) bei Gegenläufigkeit von Gruppengröße und Varianzheterogenität auf, wobei diese Unterschreitungen des Niveaus mit steigendem Parameterwert abnehmen. Eine Annäherung der Überdeckungswahrscheinlichkeit an das nominale Niveau in Abhängigkeit von N ist jedoch bei Parameterwerten zwischen $.5$ und $.714$ hier nicht festzustellen. Bei den anderen Situationen verändert sich die Überdeckungswahrscheinlichkeit von NEW-M-3 kaum in Abhängigkeit von der Stichprobengröße, hingegen wird die Überdeckung mit steigendem Parameterwert bei ungleich großen Gruppen und homogenen Varianzen ($\{1:1;1:3\}$) zunehmend konservativ, und bei Gleichsinnigkeit von Disbalance und Heterogenität ist die Überdeckung bei $\theta = 0.5$ konservativ und gelingt erst mit zunehmendem Parameterwert besser.

Unter den stetigen Verteilungen zeigt das Shift-Intervall S-BM-Z-3 für den hier realisierten Parameterbereich das insgesamt gleichmäßigste Überdeckungsverhalten, auch bei Gegenläufigkeit von Disbalance und Heterogenität liegt die empirische Überdeckungswahrscheinlichkeit über alle Stichprobengrößen hinweg zwar etwas unterhalb, aber vergleichsweise nahe beim nominalen Niveau. Zudem deutet sich über alle Bedingungen hinweg mit steigendem Stichprobenumfang eine Annäherung des empirischen an das nominale Niveau an, was hier als Hinweis für die Asymptotik des Verfahrens gewertet wird.

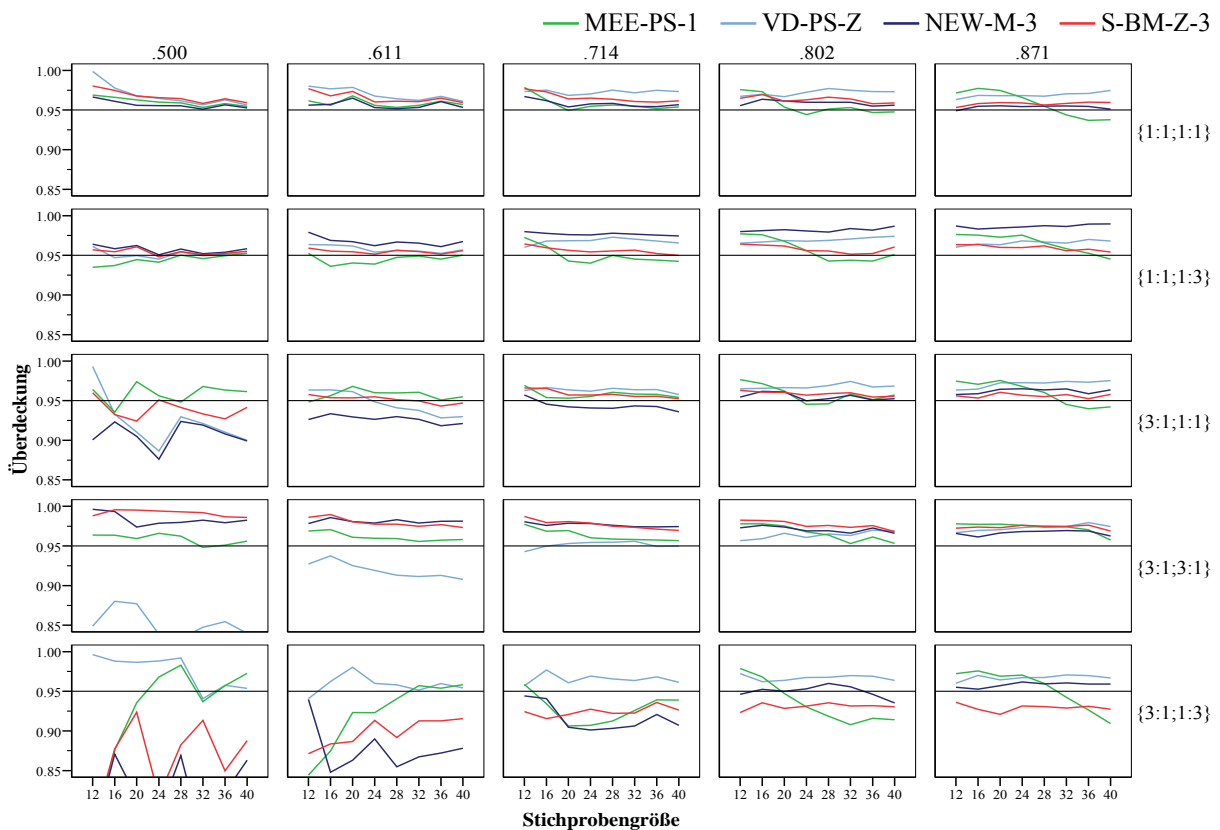


Abbildung 42. Simulierte Überdeckungswahrscheinlichkeiten ausgewählter Verfahren unter den negativen hypergeometrischen Verteilungen

Während die Ergebnisse unter den negativen hypergeometrischen Verteilungen (siehe Abbildung 42) bei den Situationen mit homogenen Varianzen insgesamt sehr ähnlich zu den Ergebnissen unter den stetigen Verteilungen sind, treten in den varianzheterogenen Bedingungen $\{3:1;1,1\}$ und $\{3:1;1,3\}$ insbesondere bei $\theta = .5$ deutlich andere Kurvenverläufe auf. Dies liegt, wie im folgenden an einem Beispiel illustriert werden soll, an der von den stetigen Modellen abweichenden Konstruktion der Simulationsbedingungen bei dem diskreten Modell, bei welchem keine lineare Transformation vorgenommen wird, sondern der Wertebereich Null bis Zehn konstant gehalten wird (vgl. Kapitel 2.1.4.2). Dieses Vorgehen erzeugt bei der hier simulierten Varianzheterogenität in einzelnen Bedingungen jedoch extreme Konstellationen.

Die deutlich ausgeprägten Sprünge in den Überdeckungsfunktionen lassen sich mit dem diskreten mehrmodalen Charakter der Stichprobenverteilungen der stochastischen Überlegenheit erklären. Abbildung 43 zeigt beispielhaft die den Simulationen zu $\theta = .5$, $N = 24$ und $\{\sigma_1:\sigma_2; n_1:n_2\} = \{3:1;1:3\}$ zugrundeliegenden negativen hypergeometrischen Verteilungen sowie die aus diesen gezogene Stichprobenverteilung von $\hat{\theta}$. Diese Stichprobenverteilung weist jeweils zu ungefähr 6% Schätzwerte von .167 und .833 auf. Ein Verfahren, welches nun für diese oder extremere Schätzwerte den Parameter $\theta = .5$ nicht überdeckt, kann die nominale Überdeckungswahrscheinlichkeit nicht einhalten; dies ist hier für NEW-M-3 und S-BM-Z-3 der Fall (vgl. Abbildung 42). Mit steigendem N ergibt sich das Phänomen, daß diese gehäuft auftretenden Werte mal zu überdeckenden Intervallen führen und mal nicht.

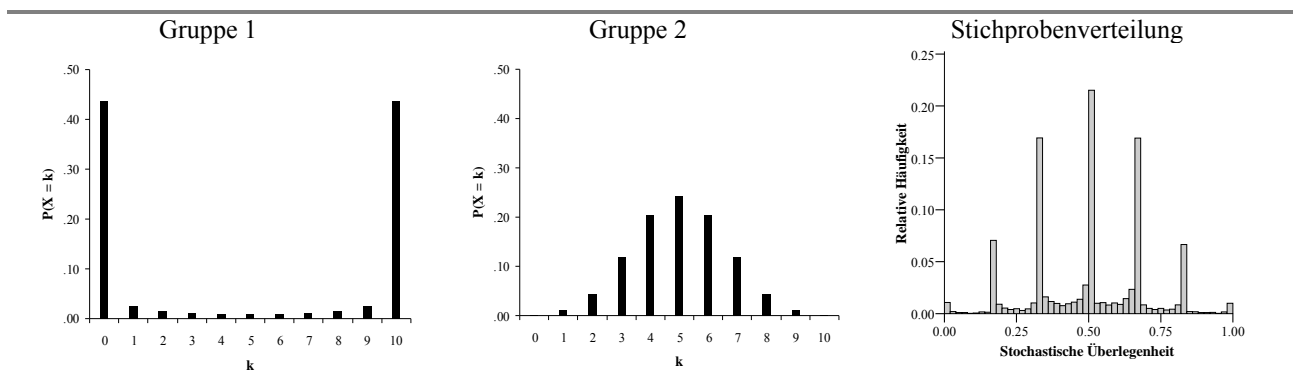


Abbildung 43. Negatives hypergeometrisches Verteilungsmodell und Stichprobenverteilung für $\theta = .500$; $\{\sigma_1:\sigma_2; n_1:n_2\} = \{3:1;1,3\}$; $N = 24$

Anders als bei den stetigen Verteilungen ist unter dem diskreten Modell kein über alle Bedingungen hinweg als robust zu bezeichnendes Verfahren auszumachen. Auch das insgesamt beste Verfahren MEE-PS-1 zeigt bei Varianzheterogenität zum Teil deutliche Unterschreitungen des Konfidenzniveaus. Abschließend sind die Verteilungen der Überdeckungswahrscheinlichkeiten und Intervallängen der vier Verfahren getrennt nach Verteilungstyp und Stichprobengröße anhand von Boxplots dargestellt (Abbildung 44). Die Überdeckung von MEE-PS-1 und S-BM-Z-3 gelingt sowohl hinsichtlich des Medians als auch in ihrer Variabilität bei größeren Stichproben besser. Hingegen ist die Variabilität der empirischen Überdeckungsrate bei VD-PS-Z und NEW-M-3 bei größeren Stichproben nicht geringer, zudem liegt bei VD-PS-Z auch der Median nicht näher am nominalen Niveau. Bezüglich der Intervallängen verdeutlicht Abbildung 44 nochmals, daß NEW-M-3 insgesamt die kürzesten Intervalle erzeugt. Unter stetigen Verteilungen zeigt VD-PS-Z insbe-

sondere bei kleinen Stichproben vergleichsweise lange Intervalle, die Verfahren MEE-PS-1 und S-BM-Z-3 zeigen hier ähnliche Verteilungen.

Zusammenfassend läßt sich folgendes sagen: Unter metrischen, kontinuierlichen Verteilungen, die hier nach dem allgemeinen Lagen-Skalen-Modell konstruiert wurden, schneidet das Shift-Intervall S-BM-Z-3 am besten ab. Für diskrete Verteilungen ist dieses jedoch nicht zu empfehlen, hier schneidet insgesamt MEE-PS-1 am besten ab. Für ordinale Variablen kann insgesamt das von Newcombe modifizierte Verfahren nach Mee (MEE-PS-2) empfohlen werden.

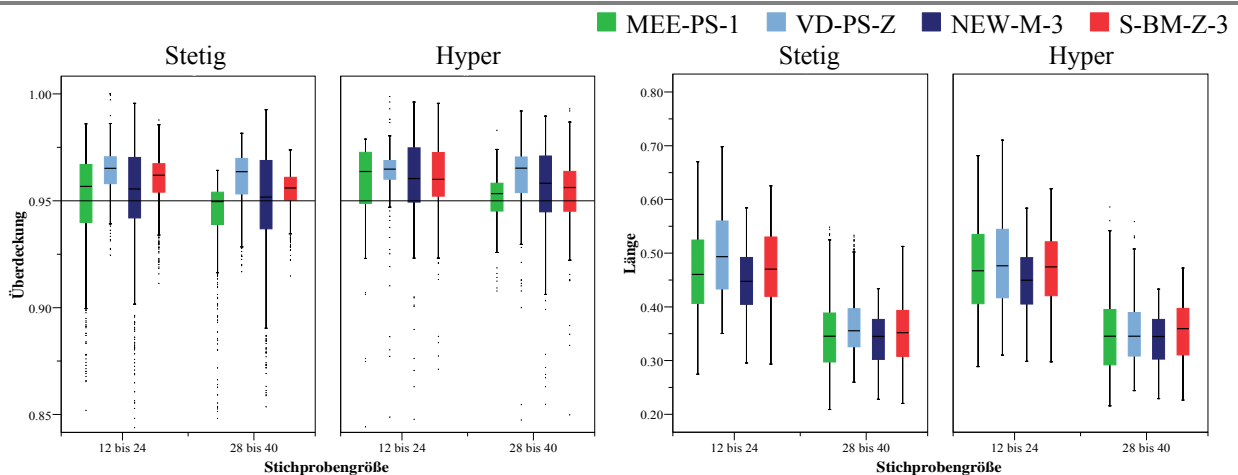


Abbildung 44. Simulierte Überdeckungswahrscheinlichkeiten und Intervalllängen ausgewählter Verfahren getrennt nach Verteilungstyp und Stichprobengröße

In einer neueren Arbeit von Ryu und Agresti (2008) wurden für geordnete kategoriale (d.s. diskrete ordinale) Daten Score-Intervalle auf Basis einer multinomialen Modellierung entwickelt. Hierbei werden die Intervallgrenzen aus der Schätzung aller $2(c - 1)$ vielen Wahrscheinlichkeitsparameter der einzelnen c Kategorien in den beiden Verteilungen berechnet. Dieses algorithmisch sehr anspruchsvolle Verfahren konnte in der vorliegenden Untersuchung nicht mehr berücksichtigt werden, verspricht aber für den hier betrachteten diskreten Fall insbesondere auch bei sehr kleinem Stichprobenumfang und vielen Bindungen hervorragende Schätzeigenschaften. In der von Ryu und Agresti berichteten Simulation lag die absolute Abweichung der empirischen Überdeckungswahrscheinlichkeit von der nominalen stets unterhalb von .02. In weiteren Studien sollte insbesondere untersucht werden, ob dieser multinomiale Ansatz auch bei stetigen Verteilungen sowie generell auch bei varianzheterogenen Verteilungen tauglich ist und wie er im Vergleich mit den Verfahren nach Mee sowie auch dem Shift-Ansatz zu beurteilen ist.

3.2 Verbundene Stichproben

Mit der stochastischen Überlegenheit im Falle verbundener Stichproben hat sich vor allem Cliff (1993, 1996a, 1996b) beschäftigt und Cliff ist m.W. auch der einzige, der analog zu den standardisierten Lagedifferenzen bei Abhängigkeit zwei Überlegenheitsmaße unterscheidet, wobei das eine die (intra-)individuelle Veränderung und das andere die Veränderung der Gruppe als Ganzes erfaßt.

3.2.1 Parameter und Punktschätzer

Für die Veränderung innerhalb der Meßwertpaare wird folgende Wahrscheinlichkeit betrachtet:

$$\theta_w = \Pr(X_{1i} > X_{2i}) + \frac{1}{2}\Pr(X_{1i} = X_{2i}),$$

die über die intraindividuellen Vergleiche innerhalb einer Stichprobe geschätzt werden kann:

$$\hat{\theta}_w = \frac{\sum_{i=1}^n U_{ii}}{n} = \frac{x}{n}.$$

Im bindungsfreien Fall ist $\hat{\theta}_w = \hat{p}$ der (Maximum-Likelihood-)Schätzer der Erfolgswahrscheinlichkeit P einer Binomialverteilung, wobei $\hat{p} = x/n$ und x die Anzahl der Erfolge oder Ereignisse bei n Wiederholungen eines Bernoulli-Experimentes sind. Unter Normalverteilungen gilt zudem folgende Beziehung zur Effektgröße des t -Tests: $\theta_w = \Phi(d_D)$. Bei diskreten Daten ist die Varianz von $\hat{\theta}_w$ auch von der Wahrscheinlichkeit abhängig, mit der Bindungen auftreten. Gemäß dem Modell des Vorzeichen-tests wird in beiden Fällen von asymptotischer Normalverteilung ausgegangen.

Die mit der stochastischen Überlegenheit θ im unabhängigen Fall identische Statistik

$$\theta_b = \Pr(X_1 > X_2) + \frac{1}{2}\Pr(X_1 = X_2)$$

kann nur unter Auslassung der individuellen Veränderung erwartungstreu geschätzt werden über

$$\hat{\theta}_b = \frac{\sum_{i=1}^n \sum_{j \neq i}^n U_{ij}}{n(n-1)}.$$

Während $E(\hat{\theta}_b) = \theta_b = \theta$, ist $\hat{\theta}$ bei Abhängigkeit nur ein asymptotisch erwartungstreuer Schätzer mit $E(\hat{\theta}_n) = \theta + (\theta_w - \theta)/n$ (Cliff, 1996a, S. 160). Tabelle 62 gibt für jeweils einige Werte von θ_b sowie ρ die zugehörigen Werte von θ_w unter Normalverteilungen an. Mit diesen Werten lassen sich nun die Erwartungswerte von $\hat{\theta}$ berechnen, z.B. lautet dieser für $\theta_b = .714$, $\rho = .8$ und $n = 20$: $E(\hat{\theta}_{20}) = .723$.

Tabelle 62. Parameterwerte von θ_w unter Normalverteilungen in Abhängigkeit von θ_b und der Korrelation ρ

θ_b	$\rho = -.8$	$\rho = -.4$	$\rho = 0$	$\rho = .4$	$\rho = .8$
.500	.500	.500	.500	.500	.500
.611	.583	.594	.611	.642	.736
.714	.663	.684	.714	.767	.897
.802	.736	.763	.802	.863	.971
.871	.800	.831	.871	.928	.994

3.2.2 Intervallschätzer

Da $\hat{\theta}_w$ im stetigen Fall einer Binomialverteilung folgt, liegt hierfür eine Vielzahl von Intervallschätzern vor (Agresti & Coull, 1998; Newcombe, 1998; Vollset, 1993; Wilcox, 2005; Zhou, Li & Yang, 2008). Hingegen wurde für die analog zum unabhängigen Fall definierte stochastische Überlegenheit θ_b ausschließlich von Cliff ein Vorschlag für die Konstruktion eines Konfidenzintervalls gemacht.

3.2.2.1 Stochastische Überlegenheit der intraindividuellen Veränderung

Im folgenden werden zunächst die aktuell empfohlenen Intervallschätzer für den Anteilswert einer Binomialverteilung dargestellt. Da bei der Schätzung von $\hat{\theta}_w$ Bindungen berücksichtigt werden – und $\hat{\theta}_w$ also nicht mehr einer Binomialverteilung folgt –, soll auch die Tauglichkeit der etablierten Verfahren im Rahmen der Simulationen beurteilt werden. Vollset (1993) empfiehlt aufgrund seiner Analyse von 13 Verfahren hinsichtlich Überdeckung, Intervallbreite und relativem Fehler folgende vier Verfahren: (1) Clopper-Pearson, (2) kontinuieritätskorrigiertes Score-Intervall nach Blyth und Still, (3) MID-P und (4) das unkorrigierte Score-Intervall, wobei die Reihenfolge der Überdeckung der Verfahren entspricht, vom konservativsten zum progressivsten. Die generelle Eignung der Score-Intervalle wird auch durch die Simulationsstudien von Newcombe (1998) sowie Zhou, Li und Yang (2008) bestätigt. Als weitere Verfahren wurden ferner das adjustierte Wald-Intervall von Agresti und Coull (1998) ausgewählt, welches sich durch eine sehr robuste Überdeckung auch bei extremen Parameterwerten auszeichnet (Agresti & Caffo, 2000), sowie die Approximation der Binomialverteilung nach Pratt (1968), die u.a. von Wilcox (2005) empfohlen wird. Verfahren zur Konstruktion extremer Intervalle (Bunke, 1959; Collani & Dräger, 2001), bei welchen das Konfidenzniveau analog zu Clopper-Pearson nie unterschritten wird, zugleich aber eine Minimierung der mittleren Intervalllänge angestrebt wird, wurden hier nicht berücksichtigt.

3.2.2.1.1 Stetige Verteilungen

Im folgenden wird weitgehend die originale Notation der Verfahren übernommen, die Anzahl der Erfolge wird mit x notiert, die Erfolgswahrscheinlichkeit mit P und ihr Schätzer mit $\hat{p} = x/n$.

3.2.2.1.1.1 Wald-Intervall

Da die Teststatistik $(\hat{p} - P) / \sqrt{\hat{p}(1 - \hat{p})/n}$ mit $n \rightarrow \infty$ in die Standardnormalverteilung übergeht (Rohatgi, 1984), kann über eine Invertierung ein approximatives Konfidenzintervall für den Wahrscheinlichkeitswert einer Binomialverteilung konstruiert werden:

$$\hat{p} \pm z_{\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n}.$$

Dieses sogenannte Wald-Intervall hat sich jedoch selbst bei größeren Stichproben wiederholt als untauglich erwiesen (Agresti & Coull, 1998; Vollset, 1993) und wird in dieser ursprünglichen Form hier nicht weiterverfolgt. Wie Agresti und Coull jedoch zeigen konnten, führt eine einfache Modifikation zu einem der aktuell besten Verfahren (siehe Kapitel 3.2.2.1.1.4).

3.2.2.1.1.2 Score-Intervall nach Wilson

Das Score-Intervall nach Wilson (1927) resultiert ebenfalls aus der Invertierung einer approximativ normalverteilten Teststatistik, wobei hier aber der Standardfehler nicht geschätzt, sondern aus der Nullhypothese übernommen wird. Als Konfidenzgrenzen werden die beiden Parameterwerte gewählt, für die $|(\hat{p} - P)/\sqrt{p(1-p)/n}| = z_{1-\alpha/2}$ gilt. Quadrieren und Umformen dieser Gleichung führt zu folgendem Intervall:

$$\frac{x + z_{1-\alpha/2}^2/2 \pm z_{1-\alpha/2} \sqrt{x - x^2/n + z_{1-\alpha/2}^2/4}}{n + z_{1-\alpha/2}^2}.$$

3.2.2.1.1.3 Score-Intervall mit Kontinuitätskorrektur nach Blyth und Still

Blyth und Still (1983) konnten zeigen, daß Unterschreitungen der nominalen Überdeckungswahrscheinlichkeit, die bei dem Score-Intervall bei Parameterwerten nahe Null oder Eins auftreten, durch die Verwendung der Kontinuitätskorrektur weitgehend vermieden werden können. Die Grenzen des so korrigierten Score-Intervalls lauten:

$$\frac{(x \pm .5) + z_{1-\alpha/2}^2/2 \pm z_{1-\alpha/2} \sqrt{(x \pm .5) - (x \pm .5)^2/n + z_{1-\alpha/2}^2/4}}{n + z_{1-\alpha/2}^2}.$$

3.2.2.1.1.4 Adjustiertes Wald-Intervall

Das von Agresti und Coull (1998) vorgeschlagene Verfahren ist der Form nach ein Wald-Intervall, welches dadurch adjustiert wird, daß es um den Mittelpunkt des Score-Intervalls \tilde{p} gelegt wird:

$$\tilde{p} \pm z_{\alpha/2} \sqrt{\tilde{p}(1-\tilde{p})/\tilde{n}}, \text{ wobei}$$

$$\tilde{p} = (n\hat{p} + z_{\alpha/2}^2/2)/(n + z_{\alpha/2}^2) \text{ und } \tilde{n} = n + z_{\alpha/2}^2.$$

Dieses relativ einfache Verfahren gehört zu den besten approximativen Methoden zur Intervallschätzung von P (Sauro & Lewis, 2005).

3.2.2.1.1.5 Exaktes Intervall nach Clopper und Pearson

Bei dem exakten Intervall nach Clopper und Pearson (1934) wird vergleichbar mit dem Score-Intervall für jede Grenze der Parameter bestimmt, für welchen der Test bei einem beobachteten Wert gerade noch die Nullhypothese beibehält, wobei hier der exakte Test auf Basis der Binomialverteilung herangezogen wird:

$$UG = \left\{ P : \sum_{t=0}^{x-1} \binom{n}{t} P^t (1-P)^{n-t} = 1 - \alpha/2 \right\},$$

$$OG = \left\{ P : \sum_{t=0}^x \binom{n}{t} P^t (1-P)^{n-t} = \alpha/2 \right\}.$$

Die Berechnung kann wie dargestellt anhand der Binomialverteilung oder auch über Quantile der F -Verteilung vorgenommen werden (siehe Kapitel 3.2.2.1.2.2). Der diesem Clopper-Pearson-Intervall zugrundeliegende Binomialtest nimmt die Punktwahrscheinlichkeit des beobachteten Wertes mit vollem Gewicht in die Berechnung des p -Wertes (des Tests) auf. Bei diesem Vorgehen (im folgenden mit MAX bezeichnet) ist garantiert, daß die Überdeckungswahrscheinlichkeit den nominalen Wert nicht unterschreiten kann. Aufgrund der Diskretheit der Binomialverteilung ist aber die mittlere Konservativität dieses Verfahrens insbesondere bei kleinem n recht hoch (Vollset, 1993). Um die Konservativität abzuschwächen, kann die Wahrscheinlichkeit des beobachteten Wertes mit einem Gewicht kleiner 1 berücksichtigt werden, wobei in der Regel ein Gewicht von .5 gewählt wird. Die Grenzen dieses im folgenden MID-P bezeichneten Verfahrens lauten:

$$UG = \left\{ P : \frac{1}{2} \binom{n}{x} P^x (1-P)^{n-x} + \sum_{t=0}^{x-1} \binom{n}{t} P^t (1-P)^{n-t} = 1 - \alpha/2 \right\},$$

$$OG = \left\{ P : \frac{1}{2} \binom{n}{x} P^x (1-P)^{n-x} + \sum_{t=0}^{x-1} \binom{n}{t} P^t (1-P)^{n-t} = \alpha/2 \right\}.$$

Für das MID-P-Intervall läßt sich keine geschlossene Formel angeben (Berry & Armitage, 1995).

3.2.2.1.1.6 Approximation der Binomial-Verteilung nach Pratt

Da für die Berechnung des exakten Intervalls die Kenntnis der jeweiligen Binomial- bzw. F -Verteilung gegeben sein muß, wurden zahlreiche Approximationen der Binomialverteilung entwickelt (Blyth, 1986). Von Pratt (1968) stammt die folgende von Blyth sowie von Wilcox (2005) empfohlene Approximation über die Standardnormalverteilung:

Falls $x = 0$: $UG = 0$; $OG = 1 - (\alpha/2)^{1/n}$;

falls $1 \leq x \leq n-1$:

$$UG = 1 - \left[1 + \left(\frac{n-x+1}{x} \right)^2 \left(\frac{81x(n-x+1) - 9n - 8 - 3z_{1-\alpha/2} \sqrt{9x(n-x+1)(9n+5 - z_{1-\alpha/2}^2) + n+1}}{81(n-x+1)^2 - 9(n-x+1)(2 + z_{1-\alpha/2}^2) + 1} \right)^3 \right]^{-1};$$

$$OG = \left[1 + \left(\frac{x+1}{n-x} \right)^2 \left(\frac{81(x+1)(n-x) - 9n - 8 - 3z_{1-\alpha/2} \sqrt{9(x+1)(n-x)(9n+5 - z_{1-\alpha/2}^2) + n+1}}{81(x+1)^2 - 9(x+1)(2 + z_{1-\alpha/2}^2) + 1} \right)^3 \right]^{-1};$$

falls $x = n$: $UG = (\alpha/2)^{1/n}$; $OG = 1$.

3.2.2.1.2 Diskrete Verteilungen

Im Zusammenhang des Vorzeichentests werden echte Bindungen von solchen unterschieden, die auf unpräzise Messungen oder Rundungen zurückgehen (Wittkowski, 1998). Im ersten Fall geben Bindungen die tatsächliche Gleichheit von Meßwerten an. Die Testung der Hypothese $H_0: P = .5$ erfolgt dann anhand des üblichen Vorzeichentests T_{ANU} („asymptotic uniformly most powerful nonrandomized test“), bei dem die Stichprobe effektiv um die gebundenen Werte verkleinert wird (Putter, 1955):

$$T_{ANU} = \frac{(N_+ + P_H N_0)/n - P_H}{\frac{1}{n} \sqrt{P_H(1 - P_H)(n - N_0)}} \xrightarrow{L} N(0,1),$$

wobei N_+ die Anzahl der positiven, N_- die Anzahl der negativen Vorzeichen, N_0 die Anzahl der Bindungen, $n = N_+ + N_0 + N_-$ und P_H die Erfolgswahrscheinlichkeit unter der Nullhypothese bezeichnen. Im zweiten Fall gehen Bindungen auf eigentlich ungleich große Werte zurück, so daß sich eine Reduktion des Stichprobenumfangs verbietet. Vielmehr ist hier ein Kalkül anzuwenden, welches gebundenen Werten ein Vorzeichen zuweist. Ein gebräuchliches Verfahren ist der „nonrandomized unconditional exact test“:

$$T_{NUE} = \frac{(N_+ + P_H N_0)/n - P_H}{\sqrt{P_H(1 - P_H)}/n} \xrightarrow{L} N(0,1).$$

Bei $H_0: P_H = .5$ und einer bindungsfreien Stichprobe mit $N_+ = 10$, $N_0 = 0$ und $N_- = 0$ lauten beide asymptotisch standardnormalverteilten Statistiken $T_{NUE} = T_{ANU} = 3.162$. Bei $N_+ = 10$, $N_0 = 90$ und $N_- = 0$ ist weiterhin $T_{ANU} = 3.162$, hingegen lautet $T_{NUE} = 1$.

Die bisher dargestellten approximativen Verfahren für stetige Variablen nutzen den Standardfehler des T_{NUE} . Da in den vorliegenden Untersuchungen jedoch mit den hypergeometrischen Verteilungen echte diskrete und keine diskretisierten Verteilungen betrachtet werden sollen, führt dieser Standardfehler mit zunehmender Wahrscheinlichkeit von Bindungen zu einer Überschätzung der Varianz. Für die Schätzung des Standardfehlers von $\hat{\theta}_w$ im diskreten Fall bietet sich dem T_{ANU} entsprechend folgende Statistik an (vgl. Putter, 1955, S. 372):

$$\begin{aligned} S_{\hat{\theta}_w} &= \sqrt{[N_+(1 - \hat{\theta}_w)^2 + N_0(.5 - \hat{\theta}_w)^2 + N_-(0 - \hat{\theta}_w)^2]/(n-1)} \\ &= \sqrt{[\hat{\theta}_w(1 - \hat{\theta}_w) - N_0/(4n)]/(n-1)}. \end{aligned}$$

Dieser Schätzer kann im Zusammenhang der beschriebenen Verfahren am einfachsten mit dem adjustierten Intervall nach Agresti und Coull (1998) verknüpft werden. Eine Berücksichtigung der Bindungen ist mit den anderen approximativen Verfahren nicht unmittelbar möglich. Bei den Score-Intervallen müßte z.B. eine Anpassung der Wahrscheinlichkeit von Bindungen an die gesuchten Parameter erfolgen. Von der Verteilung der Bindungsanzahl ist jedoch lediglich bekannt, daß der maximale Wert $N_0 = n$ nur für $\theta_w = .5$ möglich ist und daß $N_0 = 0$ ist, wenn $\theta_w = 0$ oder 1 ist. Der genaue Zusammenhang zwischen Parameterwert und erwarteter Anzahl von Bindungen ist unbekannt.

3.2.2.1.2.1 Adjustiertes Wald-Intervall

Gebundene Werte lassen sich mit dem adjustierten Wald-Intervall nun wie folgt berücksichtigen:

$$\tilde{\theta}_w \pm z_{\alpha/2} \sqrt{[\tilde{\theta}_w(1-\tilde{\theta}_w) - \tilde{N}_0/(4\tilde{n})]/\tilde{n}},$$

wobei $\tilde{N}_0 = \tilde{n}N_0/n$. Im Falle stetiger Daten führt dieses Verfahren zu denselben Grenzen wie das herkömmliche nach Agresti und Coull (1998), im Falle gebundener Werte wird mit einer entsprechend der Häufigkeit gebundener Werte reduzierten Varianz gerechnet, so daß die Grenzen näher an das Zentrum des Intervalls rücken.

3.2.2.1.2.2 MAX und MID-P

Das Clopper-Pearson-Intervall (MAX) und das MID-P-Intervall sind nur für natürliche Zahlenwerte von x definiert. Um nun auch für reelle Werte von x Intervalle zu konstruieren, wird das MAX-Intervall über die F -Verteilung berechnet (Rinne, 2003):

$$\frac{x}{x + (n-x+1)F_{v_1^*;v_2^*;1-\alpha/2}} \leq P \leq \frac{(x+1)F_{v_1;v_2;1-\alpha/2}}{n-x+(x+1)F_{v_1;v_2;1-\alpha/2}},$$

wobei $v_1^* = 2(n-x+1)$, $v_2^* = 2x$, $v_1 = 2(x+1)$ und $v_2 = 2(n-x)$. Sofern x keine natürliche Zahl ist, wird das MID-P-Intervall nach folgender Regel berechnet:

$$UG = \left\{ P : \sum_{t=0}^{[x]} \binom{n}{t} P^t (1-P)^{n-t} = 1 - \alpha/2 \right\}, \quad OG = \left\{ P : \sum_{t=0}^{[x]} \binom{n}{t} P^t (1-P)^{n-t} = \alpha/2 \right\},$$

wobei $[x]$ die nach unten abgerundete Anzahl der Erfolge ist.

3.2.2.1.2.3 Schätzer nach Cliff

Ein in der psychologischen Literatur vorgestelltes Verfahren zur Konstruktion eines Intervalls für θ_w stammt von Cliff (1993, 1996a, 1996b), wobei sich dieses im Original wiederum auf die sogenannte Dominanzstatistik $\delta_w = 2\theta_w - 1$ bezieht und hinsichtlich der Varianzschätzung äquivalent ist mit $S_{\hat{\theta}_w}$ des Vorzeichentests bei Bindungen (siehe Kapitel 3.2.2.1.2):

$$S_{D_w}^2 = \frac{\sum_{i=1}^n (D_{ii} - D_w)^2}{n(n-1)}, \text{ wobei}$$

$$D_w = \frac{\sum_{i=1}^n D_{ii}}{n} \text{ und } D_{ii} = \text{sign}(x_{1i} - x_{2i}).$$

Wie bei unabhängigen Gruppen kann auch dieser Varianzschätzer wiederum mit verschiedenen Schätzstatistiken – Wald, Logit oder Pseudo-Score (vgl. Kapitel 3.1.3) – verknüpft werden. Als kritischer Wert wird hier stets $z_{1-\alpha/2}$ eingesetzt.

3.2.2.1.3 Datenbeispiel

Die Berechnung der einzelnen Verfahren sei am Datenbeispiel aus Tabelle 63 illustriert. Da $x = 4.5$ ist, wurden das MAX- und das MID-P-Intervall nach den in Kapitel 3.2.2.1.2.2 angegebenen Formeln bestimmt.

Tabelle 63. 95%-Konfidenzintervalle für θ_w

(a)			(b)		
x_1	x_2	U_{ii}		UG	OG
23	21	1	Wilson	.463	.989
28	28	.5	Blyth & Still	.376	1.000
28	24	1	Agresti	.432	1.020
32	27	1	Agresti(No)	.472	.980
37	32	1	MAX	.371	1.000
$x = 4.5$			MID-P	.478	1.000
			Pratt	.371	1.000
			Cliff-PS	.417	.991

3.2.2.2 Stochastische Überlegenheit zwischen den Gruppen

Für die Intervallschätzung von θ_b hat bisher nur Cliff (1996a) ein Verfahren vorgestellt. Neben diesem werden Möglichkeiten anhand des t -Tests sowie des Shift-Verfahrens aufgezeigt.

3.2.2.2.1 Schätzer nach Cliff

Tabelle 64 gibt den Varianzschätzer von δ_b (Cliff, 1996a) sowie die Umrechnung in den Wahrscheinlichkeitswert θ_b an, Tabelle 65 zeigt die Berechnung an einem Datenbeispiel. Hierbei wird deutlicher, daß in diese nach Cliff erwartungstreue Schätzung der Varianz von δ_b zweierlei Kovarianzen eingehen, zum einen zwischen den Randverteilungen $D_{i\cdot}^*$ und $D_{\cdot j}^*$, zum anderen zwischen den symmetrisch um die Hauptdiagonale angeordneten Werte der Kontingenzmatrix D_{ij}^* .

Tabelle 64. Varianzschätzer von θ_b nach Cliff (1996a)

$$D_b = \frac{\sum_{i=1}^n \sum_{j \neq i}^n D_{ij}}{n(n-1)};$$

$$S_{D_b}^2 = \frac{(n-1)^2 \left(\sum_{i=1}^n D_{i\cdot}^{*2} + \sum_{j=1}^n D_{\cdot j}^{*2} + 2 \sum_{i=1}^n \sum_{j \neq i}^n D_{i\cdot}^* D_{\cdot j}^* \right) - \sum_{i=1}^n \sum_{j \neq i}^n D_{ij}^{*2} - \sum_{i=1}^n \sum_{j \neq i}^n D_{ij}^* D_{ji}^*}{n(n-1)(n-2)(n-3)}, \text{ wobei}$$

$$D_{ij}^* = D_{ij} - D_b, \quad D_{i\cdot}^* = \frac{1}{n-1} \sum_{j \neq i}^n D_{ij}^*, \quad D_{\cdot j}^* = \frac{1}{n-1} \sum_{i \neq j}^n D_{ij}^*;$$

$$\hat{\theta}_b = \frac{D_b + 1}{2}, \quad S_{\hat{\theta}_b} = \frac{S_{D_b}}{2}.$$

Wie für den unabhängigen Fall bereits demonstriert, läßt sich auch dieser Varianzschätzer mit den verschiedenen Schätzmethoden (Wald-, Logit- und Pseudo-Score-Formel) verknüpfen, wobei Cliff jedoch das Wald-Intervall empfiehlt. Als Begründung führt Cliff an, daß der Pseudo-Score-Schätzer in Simulationen deutlich geringere Teststärke aufgewiesen habe als das Wald-Intervall bzw. der Wald-Test. Diese Argumentation scheint jedoch für die Wahl eines Intervallschätzers daneben zu greifen, denn hierbei steht die Einhaltung der nominalen Überdeckungswahrscheinlichkeit im Vordergrund, welche für das Wald-Intervall jedoch nicht zu erwarten ist. Bei der Schätzung eines Parameters ist die Stärke des hierzu äquivalenten Tests der Hypothese $H_0: \theta_b = .5$ zunächst nebensächlich. Erst wenn mehrere robuste Schätzer zur Auswahl stehen, kann in einem weiteren Schritt die Entscheidung für das Verfahren mit den kürzeren Intervallen, der gewünschten Verteilung des nominalen Fehlers auf die Intervallseiten oder auch der größeren Teststärke getroffen werden.

Tabelle 65. Intervallschätzung für θ_b nach Cliff (1996a)

D_{ij}		$D_{ij}^* = D_{ij} - D_b$												
		\mathbf{x}_2												
		21	28	24	27	32								
\mathbf{x}_1	23	1	-1	-1	-1	-1	23		-1.3	-1.3	-1.3	-1.3	-1.3	-1.30
	28	1	0	1	1	-1	28	0.7		0.7	0.7	-1.3	0.20	
	28	1	0	1	1	-1	28	0.7	-0.3		0.7	-1.3	-0.05	
	32	1	1	1	1	0	32	0.7	0.7	0.7		-0.3	0.45	
	37	1	1	1	1	1	37	0.7	0.7	0.7	0.7		0.70	
							$D_{.j}^*$	0.70	-0.05	0.20	0.20	-1.05		

$$D_b = \frac{6}{20} = .3;$$

$$S_{D_b}^2 = \frac{(5-1)^2 [2.425 + 1.675 + 2(-1.575)] - 16.2 - (-9.8)}{5(5-1)(5-2)(5-3)} = 0.07\bar{3}.$$

Der Schätzer der stochastischen Überlegenheit lautet also $\hat{\theta}_b = .65$, der Standardfehler ist $S_{\hat{\theta}_b} = 0.135$ und das auf der Pseudo-Score-Formel basierende 95%-Konfidenzintervall mit kritischem z -Wert lautet: [.380; .849].

3.2.2.2.2 t-Tests auf Basis von Rängen

Auch im abhängigen Fall bietet sich mit dem über Ränge berechneten t -Test die Möglichkeit zur Konstruktion eines Intervallschätzers für θ_b . Dieser Weg soll hier trotz der theoretischen Mängel, die durch die fehlende Unabhängigkeit zwischen Zähler und Nenner von den herkömmlichen, über Ränge berechneten t -Statistiken ausgehen, verfolgt werden, da der Rang- t -Test bei verbundenen Stichproben zumindest zur Prüfung der Nullhypothese $\theta = .5$ sehr gute Eigenschaften hat (Munzel, 2006).

Für die Berechnung des Standardfehlers wird neben der herkömmlichen Vorgehensweise eine Variante der Rangvergabe vorgeschlagen. Hierbei ergibt sich der Rang eines Wertes unter Auslassung des jeweils gepaarten Wertes aus dem Vergleich mit den $2n - 1$ von diesem unabhängigen Werten. Anders als bei der üblichen Rangvergabe kann nun $\hat{\theta}_b$ aus den Mittelwerten der so vergebenen Ränge berechnet werden. Das genaue Vorgehen sei am bereits genannten Datenbeispiel demonstriert (Tabelle 66).

Tabelle 66. Intervallschätzung für θ_b mit modifiziertem Rang- t -Test

\mathbf{x}_1	\mathbf{x}_2	R_1^*	R_2^*
23	21	1.0	1.0
28	28	5.5	5.5
28	24	5.0	3.0
32	27	7.5	4.0
37	32	9.0	8.5

$$\hat{\theta}_b = \frac{\bar{R}_1^* - \frac{n+1}{2}}{n-1} = \frac{5.6 - 3}{4} = .65$$

$$S_{\bar{R}_1^* - \bar{R}_2^*} = \sqrt{\frac{S_{R_1^*}^2 + S_{R_2^*}^2 - 2S_{R_1^*, R_2^*}}{n}} = \sqrt{\frac{9.175 + 7.925 - 2 \cdot 7.388}{5}} = 0.682$$

Da für die Beziehung zwischen Mittelwertsdifferenz und stochastischer Überlegenheit hier

$$\bar{R}_1^* - \bar{R}_2^* = 2(n-1)\hat{\theta}_b + 1 - n \text{ gilt, wird als Standardfehler berechnet:}$$

$$S_{\hat{\theta}_b} = \frac{S_{\bar{R}_1^* - \bar{R}_2^*}}{2(n-1)} = \frac{0.682}{8} = 0.085.$$

Das auf der t -Verteilung basierende 95%-Konfidenzintervall mit $t_{4,0.975} = 2.776$ ist [.413; .887].

3.2.2.2.3 Shift-Konfidenzintervall

Ganz analog zum Vorgehen bei unabhängigen Stichproben können auch hier die beschriebenen Standardfehlerschätzer zur Berechnung von Shift-Intervallen eingesetzt werden. Tabelle 67 zeigt für das Datenbeispiel $\mathbf{x}_1 = (23, 28, 28, 32, 37)'$ und $\mathbf{x}_2 = (21, 28, 24, 27, 32)'$ die 26 möglichen verschiedenen Verschiebungen sowie die zugeordneten Intervalle auf Basis des modifizierten Rang- t -Standardfehlers. Die drei in Kapitel 3.1.3.7 beschriebenen Varianten zur Berechnung eines Shift-Intervalls lauten hier nach Variante 1 [.600; .750], nach Variante 2 [.500; .900] und nach Variante 3 [.503; .888].

Wie sich bereits in – hier nicht näher ausgeführten – Voruntersuchungen ergab, kann bei sehr kleinen Stichproben oder vielen Bindungen das Problem auftreten, daß die Grenzen der um die verschobenen Werte $\tilde{\theta}_b$ gelegten Intervalle nicht monoton mit $\tilde{\theta}_b$ steigen bzw. fallen. In diesen Fällen wird der weiter von $\hat{\theta}_b$ entfernt liegende Wert $\tilde{\theta}_b$ gewählt, dessen zugehöriges Intervall die jeweilige Bedingung von Variante 1 oder 2 erfüllt. Obwohl diese – jedenfalls bei kleinen Stichproben – mangelnde Monotonie der ausgewählten Statistiken zur Wahl der Intervallgrenzen das Shift-Intervall eigentlich als möglichen Schätzer für θ_b bereits verbietet, waren auch im abhängigen Fall die ersten Ergebnisse ermutigend, so daß dieses Verfahren dennoch untersucht wurde. Für eine Weiterentwicklung des Shift-Intervalls wäre aber die Verknüpfung mit anderen Standardfehlerschätzern, bei denen das beschriebene Problem nicht auftreten kann, notwendig.

Tabelle 67. Indizes der beobachteten Werte ($i = \text{Gruppe 1}, j = \text{Gruppe 2}$), paarweise Differenzen Δ_{ij} , stochastische Überlegenheit $\tilde{\theta}_b$, Standardfehler des modifizierten Rang- t -Tests sowie zugeordnete Konfidenzintervalle für $\mathbf{x}_1 - \Delta_{\min} + \varepsilon, \mathbf{x}_2$ (Index 1) und $\mathbf{x}_1 - \Delta_{ij} - \varepsilon, \mathbf{x}_2$ (Indizes 2 bis 26)

Index	i	j	Δ_{ij}	$\tilde{\theta}_b$	S_{θ}^2	UG	OG
1	1	5	-9	1.00	0.052	.855	1.145
2	1	5	-9	.95	0.061	.782	1.118
3	1	2	-5	.90	0.085	.663	1.137
4	1	4	-4	.75	0.091	.499	1.001
5	2	5	-4	.75	0.091	.499	1.001
6	3	5	-4	.75	0.091	.499	1.001
7	1	3	-1	.70	0.091	.446	.954
8	2	2	0	.60	0.081	.376	.824
9	3	2	0	.60	0.081	.376	.824
10	4	5	0	.60	0.081	.376	.824
11	2	4	1	.50	0.052	.355	.645
12	3	4	1	.50	0.052	.355	.645
13	1	1	2	.50	0.052	.355	.645
14	2	3	4	.40	0.085	.163	.637
15	3	3	4	.40	0.085	.163	.637
16	4	2	4	.40	0.085	.163	.637
17	4	4	5	.40	0.085	.163	.637
18	5	5	5	.40	0.085	.163	.637
19	2	1	7	.30	0.103	.013	.587
20	3	1	7	.30	0.103	.013	.587
21	4	3	8	.25	0.095	-.013	.513
22	4	1	9	.20	0.103	-.087	.487
23	5	4	10	.15	0.085	-.087	.387
24	5	2	11	.10	0.076	-.110	.310
25	5	3	13	.05	0.061	-.118	.218
26	5	1	16	.00	0.052	-.145	.145

3.2.3 Notation der ausgewählten Statistiken

Die Schätzer für θ_w werden im folgenden nach ihren Autoren benannt (siehe Kapitel 3.2.2.1); die dargestellten Ansätze zur Schätzung von θ_b wurden zu 21 Intervallschätzern kombiniert, deren Notation in Tabelle 68 aufgelistet ist.

Tabelle 68. Intervallschätzer der Simulationsstudie zur stochastischen Überlegenheit bei verbundenen Gruppen

Autor/Name	Formel/Verfahren	Notation
Cliff (1996a)	Wald-Intervall mit z-Wert	CLI-WALD-Z
	Wald-Intervall mit t-Wert	CLI-WALD-T
	Pseudo-Score-Schätzer mit z-Wert	CLI-PS-Z
	Pseudo-Score-Schätzer mit t-Wert	CLI-PS-T
	Logit-Wald mit z-Wert	CLI-LOG-Z
	Logit-Wald mit t-Wert	CLI-LOG-T

(Fortsetzung Tabelle 68)

Autor/Name	Formel/Verfahren	Notation
Rang- <i>t</i> -Test	modifizierter Rang- <i>t</i> , Wald-Intervall	MRT-WALD
	Rang- <i>t</i> , Wald-Intervall	RT-WALD
	modifizierter Rang- <i>t</i> , Pseudo-Score-Schätzer	MRT-PS
	Rang- <i>t</i> , Pseudo-Score-Schätzer	RT-PS
	modifizierter Rang- <i>t</i> , Logit-Wald	MRT-LOG
	Rang- <i>t</i> , Logit-Wald	RT-LOG
Shift-Intervall	Cliff- <i>z</i> , Variante 1	S-CLI-Z-1
	Cliff- <i>t</i> , Variante 1	S-CLI-T-1
	Cliff- <i>z</i> , Variante 2	S-CLI-Z-2
	Cliff- <i>t</i> , Variante 2	S-CLI-T-2
	Cliff- <i>z</i> , Variante 3	S-CLI-Z-3
	Cliff- <i>t</i> , Variante 3	S-CLI-T-3
	modifizierter Rang- <i>t</i> , Variante 1	S-MRT-1
	modifizierter Rang- <i>t</i> , Variante 2	S-MRT-2
	modifizierter Rang- <i>t</i> , Variante 3	S-MRT-3

3.2.4 Untersuchungsdesign

Studie 4 ist hinsichtlich der verwendeten Verteilungen, der Erzeugung der Pseudozufallszahlen, der Datenmodelle, der Robustheitskriterien und der Anzahl der Simulationsdurchgänge identisch mit Studie 2 (siehe Kapitel 2.2.4 sowie Anhang B). Der Gesamtstichprobenumfang wurde mit 15 Stufen zwischen $n = 6$ und $n = 20$ variiert (6, 7, ..., 20). Wie bei den anderen Studien wurden die Effektgröße θ_b mit fünf Stufen (.500, .611, .714, .802, .871) und das Varianzverhältnis $\sigma_1 : \sigma_2$ mit zwei Stufen (1:1 und 3:1) realisiert. Diese unabhängigen Variablen wurden nach folgendem Schema gekreuzt, so daß in Studie 4 insgesamt 3150 Simulationsbedingungen umgesetzt wurden:

$$\begin{matrix}
 \text{Verteilung} & N & \{\sigma_1 : \sigma_2; \rho\} & \theta_b \\
 \left(\begin{array}{l} \text{Normal} \\ \text{Gleich} \\ g-h-1 \\ g-h-2 \\ g-h-3 \\ g-h-4 \\ \text{Hyper} \end{array} \right) & \times \begin{pmatrix} 6 \\ 7 \\ \vdots \\ 20 \end{pmatrix} & \times \begin{pmatrix} 1:1; .0 \\ 1:1; .4 \\ 1:1; .8 \\ 3:1; .0 \\ 3:1; .4 \\ 3:1; .8 \end{pmatrix} & \times \begin{pmatrix} .500 \\ .611 \\ .714 \\ .802 \\ .871 \end{pmatrix} .
 \end{matrix}$$

3.2.5 Ergebnisse der Studie 4

3.2.5.1 Schätzer für θ_w

Die neun ausgewählten Schätzer seien zunächst für zwei Stichprobengrößen anhand ihrer Überdeckungsfunktionen betrachtet, die sich unter der Binomialverteilung ergeben. Hierbei wird für Parameterwerte zwischen Null und Eins folgende Wahrscheinlichkeit bestimmt:

$$\text{coverage}(P) = \sum_{x=0}^n I(x, P) \binom{n}{x} P^x (1-P)^{n-x} ,$$

wobei $I(x, P)$ indiziert, ob der betrachtete Intervallschätzer den Parameter P überdeckt, falls $X = x$ ist. Entsprechend kann auch die erwartete Länge einer Intervallschätzung bestimmt werden nach

$$E(\text{width}) = \sum_{x=0}^n \text{width}(x) \binom{n}{x} P^x (1-P)^{n-x}.$$

Für den stetigen Fall läßt sich die Güte aller ausgewählten Verfahren also auch ohne Simulationen beurteilen. Abbildung 45 und Abbildung 46 zeigen für $n = 10$ und $n = 60$ die Überdeckungsfunktionen ($\text{coverage}(P)$), wobei die Schrittweite der Erfolgswahrscheinlichkeit .001 beträgt; jede Abbildung basiert somit auf 1001 Werten.

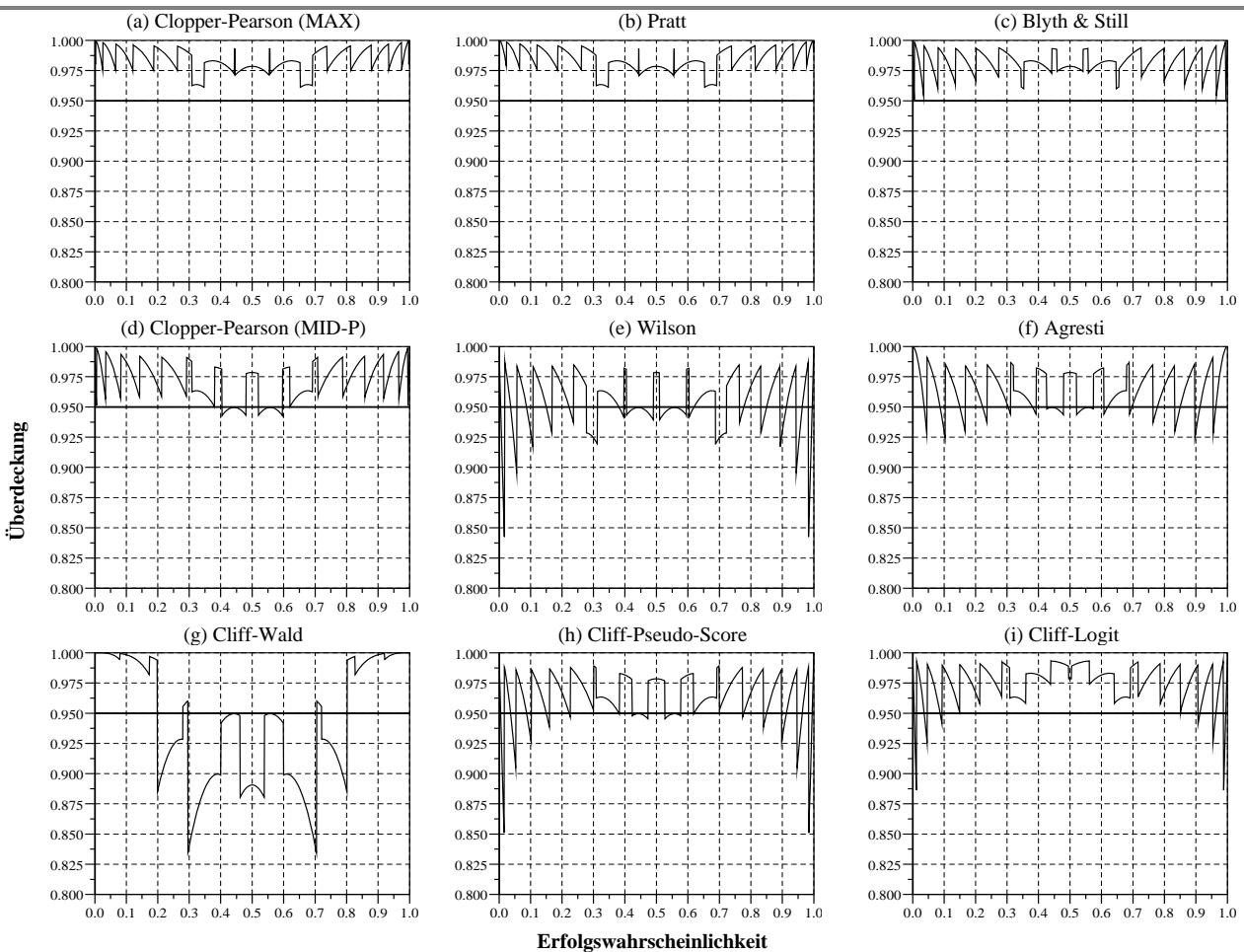


Abbildung 45. Überdeckungsfunktionen der Intervallschätzer für die Erfolgswahrscheinlichkeit bei $n = 10$

Es wird deutlich, daß das exakte Clopper-Pearson-Intervall (MAX) sowie auch die beiden Approximationen nach Pratt (1968) sowie Blyth und Still (1983) stets die Bedingung $\alpha^* \leq \alpha$ einhalten. Zudem deutet sich an, daß insbesondere die Approximation nach Pratt sehr nah an dem exakten Verfahren ist. Auch die zugehörigen Kennwerte der Überdeckungsfunktionen in Tabelle 69 und Tabelle 70 zeigen auf, daß bereits bei $n = 10$ keine relevanten Unterschiede zwischen den Überdeckungsleistungen dieser Verfahren bestehen.

Tabelle 69. Kennwerte der Überdeckungsfunktionen bei $n = 10$

	$p_{CP_{.936}}^{.966}$	$p_{CP_{.936}}^{1.000}$	$ CP - .95 $	Mittel	Min	Max	Mittel	Max
MAX	.078	<u>1.000</u>	.034	.984	.961	1.000	0.508	0.600
Pratt	.078	<u>1.000</u>	.034	.984	.961	1.000	0.510	0.602
Blyth & Still	.050	<u>1.000</u>	.032	.982	.951	1.000	0.504	0.577
MID-P	.386	<u>1.000</u>	.023	.972	.943	1.000	0.470	0.551
Wilson	<u>.539</u>	.846	<u>.017</u>	.954	.842	1.000	0.435	0.507
Agresti	.507	.970	<u>.017</u>	.965	.925	1.000	0.456	0.512
Cliff-Wald	.154	.551	.044	.943	.835	1.000	0.458	0.610
Cliff-PS	.466	.946	.019	.964	.851	1.000	0.457	0.527
Cliff-Logit	.230	.980	.027	.976	.886	1.000	0.492	0.557

Anmerkung: Unterstreichungen kennzeichnen die besten Resultate je Kriterium

Strebt man nun aber Intervalle an, deren Überdeckungswahrscheinlichkeit möglichst nahe an dem theoretischen Wert liegt und nimmt hierfür in Kauf, daß die Wahrscheinlichkeit gegebenenfalls den nominalen Wert unterschreitet und also nur $\alpha^* \approx \alpha$ erfüllt wird, so bieten sich zur Schätzung von P (bzw. θ_w) auch die Verfahren MID-P, das adjustierte Wald-Intervall nach Agresti und Coull (1998) sowie mit Einschränkungen das Score-Intervall nach Wilson (1927) und der Schätzer nach Cliff (1996a) mit der Pseudo-Score-Formel an.

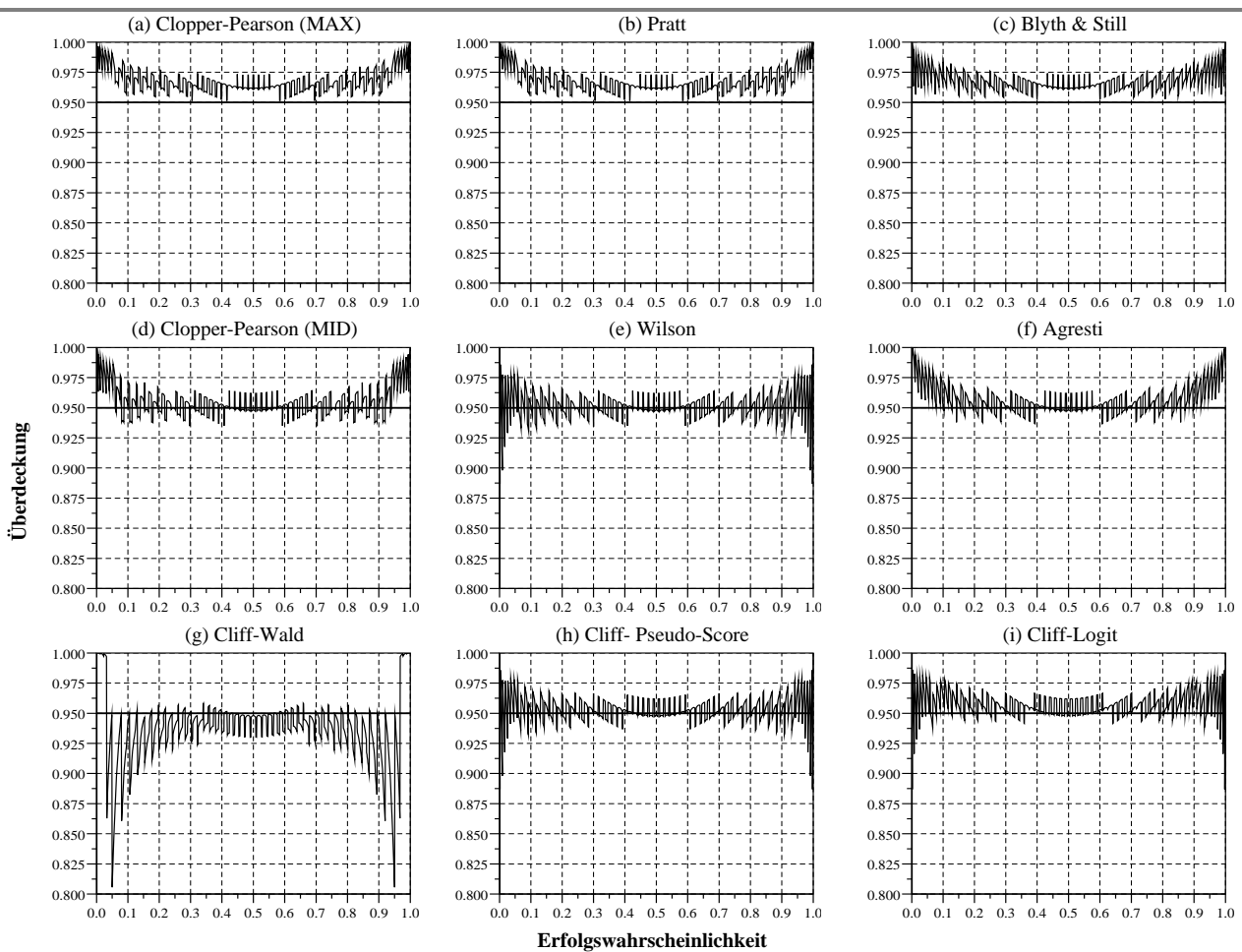


Abbildung 46. Überdeckungsfunktionen der Intervallschätzer für die Erfolgswahrscheinlichkeit bei $n = 60$

Das Score-Intervall weist zwar die geringste mittlere absolute Abweichung vom nominalen Niveau ($|\overline{CP} - .95|$) auf, unterschreitet das Niveau bei extremen Parameterwerten aber zum Teil sehr deutlich. Das adjustierte Wald-Intervall nach Agresti stellt hier einen guten Kompromiß zwischen den Eigenschaften des MID-P und des Score-Intervalls dar, da es ähnlich nah am nominalen Niveau liegt, aber geringere Abweichungen bei extremen Parameterwerten aufweist als das Score-Intervall. Das Verfahren Cliff-Wald, welches bis auf die Varianzkorrektur dem klassischen Wald-Intervall entspricht, unterschreitet die nominale Überdeckungswahrscheinlichkeit selbst bei einer Stichprobengröße von $N = 60$ zum Teil sehr deutlich und scheidet als Option für die Schätzung von P bzw. θ_w generell aus. Da die Variante Cliff-Logit bei mittleren Parameterwerten vergleichsweise konservativ ist und an den Rändern den genannten Verfahren gegenüber keine Vorteile bietet, fällt auch Cliff-Logit aus der weiteren Betrachtung raus.

Tabelle 70. Kennwerte der Überdeckungsfunktionen bei $n = 60$

	$pCP_{.936}^{.966}$	$pCP_{.936}^{1.000}$	$ \overline{CP} - .95 $	Mittel	Min	Max	Mittel	Max
MAX	.490	<u>1.000</u>	.018	.968	.951	1.000	0.210	0.262
Pratt	.490	<u>1.000</u>	.018	.968	.951	1.000	0.210	0.262
Blyth & Still	.452	<u>1.000</u>	.018	.968	.953	1.000	0.211	0.259
MID-P	.846	.996	.009	.955	.935	1.000	0.197	0.248
Wilson	.880	.956	<u>.007</u>	.952	.887	1.000	0.195	0.243
Agresti	.810	1.000	.010	.957	.937	1.000	0.199	0.243
Cliff-Wald	.540	.606	.019	.938	.806	1.000	0.196	0.253
Cliff-PS	<u>.884</u>	.970	.008	.953	.887	1.000	0.197	0.245
Cliff-Logit	.814	.992	.010	.958	.887	1.000	0.201	0.248

Anmerkung: Unterstreichungen kennzeichnen die besten Resultate je Kriterium

Im stetigen Fall ist eine Simulation der Überdeckungswahrscheinlichkeiten der Intervallschätzer also eigentlich nicht erforderlich. Im diskreten Fall sieht das anders aus, da die Bindungen nicht eliminiert werden, sondern jeder Gruppe zur Hälfte angerechnet werden. Bei dieser in der Psychologie üblichen Definition entfernt sich die Statistik jedoch mit zunehmender Anzahl der Bindungen von der Binomialverteilung. Zu erwarten ist, daß die exakten Verfahren mit zunehmender Wahrscheinlichkeit von Bindungen die nominale Überdeckungswahrscheinlichkeit überschreiten, da die Varianz im Vergleich zur Annahme des Binomialmodells abnimmt. Um diesen Effekt einschätzen zu können, wurden die Intervallschätzer auch unter den stetigen Verteilungen simuliert. Tabelle 71 gibt die Überdeckungswahrscheinlichkeiten und Längen der Intervallschätzungen für θ_w unter den stetigen Verteilungen wieder, Tabelle 72 zeigt die Ergebnisse für die negativen hypergeometrischen Modelle.

Vergleichbar mit den aus der Binomialverteilung abgeleiteten Ergebnissen liegen bei stetigen Verteilungen die Überdeckungswahrscheinlichkeiten des Score-Intervalls, des adjustierten Wald-Intervalls und von Cliff-PS im Mittel am nächsten am nominalen Niveau, $|\overline{CP} - .95|$ ist jeweils .018. Daß Cliff-PS hier ähnlich nah am nominalen Niveau liegt wie die beiden anderen Verfahren kann damit erklärt werden, daß die hier simulierten Parameterwerte die extremen Bereiche nicht gleichermaßen abdecken wie bei der Berechnung der Überdeckungsfunktionen (vgl. Tabelle 69 und Tabelle 70); zudem wurden hier die Ergebnisse für Stichproben von $n = 6$ bis 20 aggregiert. Erwartungsgemäß schneiden das exakte Verfahren, dessen Approximationen sowie MID-P unter den hypergeometrischen Verteilungen schlechter bzw. konservativer ab als unter den stetigen. Unter

dem diskreten Modell zeigen die beiden Verfahren Agresti(N_0) und Cliff-PS die geringsten Abweichungen vom nominalen Niveau, wobei die Robustheit ($pCP_{.936}^{.966}$) von Agresti(N_0) höher ist als die von Cliff-PS.

Tabelle 71. Überdeckungswahrscheinlichkeiten und Länge der Intervallschätzungen für θ_w unter stetigen Verteilungen

	$pCP_{.936}^{.966}$	$pCP_{.936}^{1.000}$	$ \overline{CP} - .95 $	Überdeckung					Länge	
				Mittel	Min	Max	LNCP	RNCP	Mittel	Max
MAX	.103	<u>1.000</u>	.032	.982	.950	1.000	.006	.012	0.466	0.720
Pratt	.102	<u>1.000</u>	.032	.982	.950	1.000	.006	.012	0.467	0.723
Blyth & Still	.107	<u>1.000</u>	.030	.980	.950	.999	.004	.016	0.463	0.688
MID-P	.330	.959	.023	.970	.917	.999	.009	.021	0.431	0.665
Wilson	<u>.504</u>	.846	<u>.018</u>	.955	.836	.989	.010	.035	0.402	0.593
Agresti	.460	.942	<u>.018</u>	.964	.908	1.000	.007	.029	0.421	0.604
Agresti(N_0)	.460	.942	<u>.018</u>	.964	.908	1.000	.007	.029	0.421	0.604
Cliff-Wald	.195	.363	.066	.901	.628	1.000	.082	.017	0.404	0.729
Cliff-PS	.456	.872	<u>.018</u>	.958	.846	.993	.012	.030	0.411	0.626
Cliff-Logit	.275	.978	.024	.972	.888	1.000	.005	.023	0.449	0.683

Anmerkung: Unterstreichungen kennzeichnen die besten Resultate je Kriterium

Hinsichtlich der Überdeckungssymmetrie ist festzustellen, daß keines der Verfahren insgesamt symmetrisch überdeckt und daß bis auf Cliff-Wald alle Verfahren bei den hier simulierten Parameterwerten $\geq .5$ in ähnlichem Ausmaß dazu neigen, diese auf der rechten Seite nicht zu überdecken.

Tabelle 72. Überdeckungswahrscheinlichkeiten und Länge der Intervallschätzungen für θ_w unter der negativen hypergeometrischen Verteilung

	$pCP_{.936}^{.966}$	$pCP_{.936}^{1.000}$	$ \overline{CP} - .95 $	Überdeckung					Länge	
				Mittel	Min	Max	LNCP	RNCP	Mittel	Max
MAX	.016	<u>1.000</u>	.040	.990	.953	1.000	.003	.007	0.472	0.730
Pratt	.016	<u>1.000</u>	.040	.990	.953	1.000	.003	.006	0.473	0.734
Blyth & Still	.013	<u>1.000</u>	.039	.989	.962	1.000	.003	.008	0.468	0.696
MID-P	.182	.989	.029	.979	.928	1.000	.007	.014	0.425	0.661
Wilson	.244	.987	.024	.973	.928	.998	.007	.021	0.407	0.601
Agresti	.147	.993	.028	.977	.934	1.000	.005	.018	0.425	0.610
Agresti(N_0)	<u>.744</u>	.891	<u>.013</u>	.953	.832	.990	.006	.041	0.394	0.599
Cliff-Wald	.060	.218	.057	.908	.744	1.000	.074	.018	0.376	0.721
Cliff-PS	.591	.880	.014	.955	.895	.985	.015	.030	0.375	0.621
Cliff-Logit	.380	.989	.019	.968	.918	.996	.009	.024	0.405	0.677

Anmerkung: Unterstreichungen kennzeichnen die besten Resultate je Kriterium

Zur Illustration des unterschiedlichen Verhaltens unter stetigen und diskreten Modellen zeigt Abbildung 47 die gemittelten Überdeckungswahrscheinlichkeiten der Verfahren von Agresti, Agresti(N_0) und Cliff-PS. Die grüne Linie im linken Teil der Abbildung zeigt das Verhalten von Agresti wie auch von Agresti(N_0), da diese Verfahren unter stetigen Verteilungen identische Ergebnisse liefern.

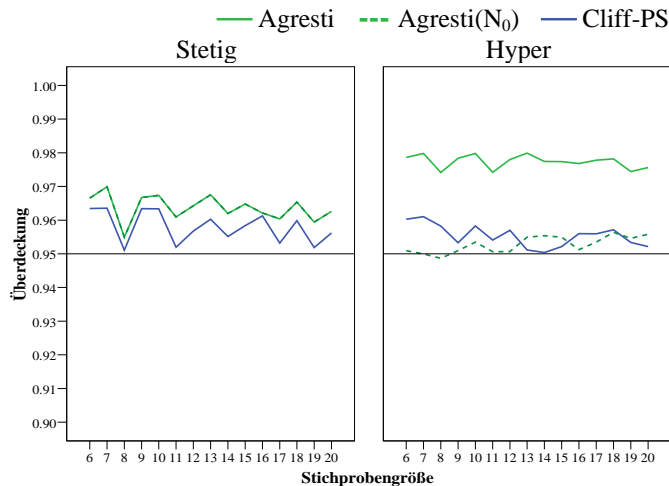


Abbildung 47. Überdeckungsverhalten des adjustierten Wald-Intervalls nach Agresti mit und ohne Verrechnung gebundener Werte im Vergleich zu Cliff-PS

Es wird deutlich, daß Agresti(N_0) und Cliff-PS, bei welchen die gebundenen Werte zur Bestimmung der Varianz berücksichtigt werden, im Gegensatz zu Agresti insgesamt unter dem negativen hypergeometrischen Modell ein ähnliches Überdeckungsverhalten zeigen wie unter den stetigen Verteilungen. Sofern also das nominale Niveau streng einzuhalten ist, bieten sich für die Schätzung von θ_w auch unter diskreten Verteilungen das Clopper-Pearson (MAX) oder eines der approximativen Verfahren (Pratt oder Blyth & Still) an. Wenn jedoch ein Intervall mit einer Überdeckungswahrscheinlichkeit nahe dem nominalen Niveau angestrebt wird, sind die Verfahren Agresti(N_0) und Cliff-PS vorzuziehen, bei welchen die Reduktion des Standardfehlers aufgrund der Bindungen berücksichtigt wird.

3.2.5.2 Schätzer für θ_b

Tabelle 73 und Tabelle 74 geben die aggregierten Überdeckungseigenschaften der Schätzer für θ_b wieder. Im Hinblick auf die Kriterien der Häufigkeit robuster Intervalle ($pCP_{.936}^{.966}$), der absoluten Abweichung vom nominalen Niveau ($|\overline{CP} - .95|$) sowie auch der größten minimalen Überdeckungswahrscheinlichkeit besitzen die Shift-Intervalle mit dem Varianzschätzer von Cliff sowie kritischen t -Werten die insgesamt besten Eigenschaften, wobei auch hier wiederum die Variante 3 (S-CLI-T-3) einen guten Kompromiß darstellt. Wie im Falle unabhängiger Stichproben schneidet das Shift-Intervall nach Methode 1 unter den stetigen Verteilungen etwas günstiger ab, wobei dieses jedoch für diskrete Verteilungen vollkommen ungeeignet ist – der $pCP_{.936}^{.966}$ -Wert von S-CLI-Z-1 ist 0 (vgl. Tabelle 74) – und daher auch hier nicht näher dargestellt wird. Diese mangelnde Eignung für diskrete Daten zeigt sich auch für S-CLI-T-3, insgesamt gelingt die Überdeckung schlechter als unter stetigen Verteilungen. Neben S-CLI-T-3 weisen die Verfahren CLI-LOG-T und CLI-PS-T die nächst größeren absoluten Abweichungen vom nominalen Niveau auf und werden im folgenden dem Shift-Intervall gegenübergestellt. Im Vergleich zur Wald-Methode führen wiederum die Pseudo-Score-Formel und die Logit-Transformation mit den beiden rangbasierten Varianzschätzern des t -Tests zu robusteren Intervallen, die Überdeckungswahrscheinlichkeiten liegen jedoch weder näher am nominalen Niveau noch fallen die Intervalle kürzer aus als bei den drei genannten Verfahren.

Tabelle 73. Überdeckung und Länge der Intervallschätzungen bei stetigen Modellen

	$pCP_{936}^{.966}$	$pCP_{936}^{1.000}$	$ CP - .95 $	Mittel	Min	Max	LNCP	RNCP	Mittel	Max
CLI-WALD-Z	.035	.035	.062	.888	.626	.945	.074	.038	0.332	0.674
CLI-WALD-T	.374	.374	.040	.911	.643	.963	.062	.026	0.373	0.799
CLI-PS-Z	.387	.387	.039	.912	.675	.962	.042	.047	0.320	0.574
CLI-PS-T	.560	.677	.025	.937	.703	.986	.030	.033	0.358	0.673
CLI-LOG-Z	.520	.523	.031	.922	.683	.972	.035	.043	0.336	0.620
CLI-LOG-T	.459	.754	.023	.947	.774	.990	.023	.030	0.381	0.740
MRT-WALD	.434	.570	.033	.928	.804	.995	.037	.035	0.354	0.755
RT-WALD	.402	.728	.027	.946	.821	.997	.031	.023	0.386	0.841
MRT-PS	.355	.706	.026	.945	.822	.990	.011	.045	0.359	0.688
RT-PS	.192	.802	.027	.961	.863	.996	.007	.031	0.387	0.724
MRT-LOG	.291	.724	.027	.950	.827	.994	.009	.041	0.391	0.837
RT-LOG	.157	.821	.029	.965	.867	.998	.006	.029	0.423	0.876
S-CLI-Z-1	.384	.384	.029	.922	.718	.964	.028	.050	0.311	0.500
S-CLI-T-1	<u>.689</u>	.847	<u>.014</u>	.951	.828	.994	.016	.033	0.352	0.603
S-CLI-Z-2	.574	.593	.023	.929	.726	.976	.026	.045	0.324	0.554
S-CLI-T-2	.542	.849	.018	.954	.804	.992	.015	.031	0.362	0.650
S-CLI-Z-3	.505	.506	.025	.926	.731	.969	.027	.047	0.318	0.532
S-CLI-T-3	.617	<u>.865</u>	.015	.954	.833	.990	.015	.031	0.357	0.627
S-MRT-1	.407	.633	.032	.934	.775	.990	.023	.043	0.340	0.664
S-MRT-2	.386	.746	.028	.946	.808	.993	.019	.035	0.358	0.723
S-MRT-3	.403	.691	.029	.941	.800	.991	.021	.038	0.349	0.693

Anmerkung: Unterstreichungen kennzeichnen die besten Resultate je Kriterium

Tabelle 74. Überdeckung und Länge der Intervallschätzungen bei negativen hypergeometrischen Verteilungen

	$pCP_{936}^{.966}$	$pCP_{936}^{1.000}$	$ CP - .95 $	Mittel	Min	Max	LNCP	RNCP	Mittel	Max
CLI-WALD-Z	.058	.058	.060	.890	.677	.954	.074	.036	0.343	0.711
CLI-WALD-T	.382	.387	.037	.914	.703	.972	.061	.025	0.385	0.832
CLI-PS-Z	.373	.400	.037	.915	.703	.985	.040	.044	0.328	0.608
CLI-PS-T	.476	.669	.025	.940	.731	.992	.029	.031	0.366	0.706
CLI-LOG-Z	.433	.513	.031	.926	.712	.991	.033	.041	0.344	0.662
CLI-LOG-T	.458	.771	.023	.950	.808	.998	.022	.028	0.390	0.784
MRT-WALD	.424	.529	.046	.912	.699	.988	.046	.041	0.346	0.749
RT-WALD	.322	.640	.038	.932	.757	.994	.039	.029	0.380	0.839
MRT-PS	.338	.647	.035	.934	.721	.989	.016	.050	0.349	0.656
RT-PS	.182	.753	.033	.953	.771	.999	.011	.036	0.378	0.710
MRT-LOG	.289	.673	.035	.940	.727	.993	.013	.047	0.380	0.812
RT-LOG	.167	.796	.033	.958	.776	1.000	.009	.033	0.412	0.854
S-CLI-Z-1	.000	.000	.210	.740	.303	.855	.095	.165	0.229	0.388
S-CLI-T-1	.000	.000	.162	.788	.451	.915	.071	.141	0.264	0.483
S-CLI-Z-2	.189	.927	.027	.971	.834	.996	.007	.022	0.421	0.673
S-CLI-T-2	.116	<u>.971</u>	.033	.980	.879	.997	.004	.016	0.456	0.735
S-CLI-Z-3	.442	.442	.033	.917	.721	.956	.025	.058	0.321	0.515
S-CLI-T-3	<u>.629</u>	.778	<u>.020</u>	.943	.814	.979	.015	.042	0.358	0.600
S-MRT-1	.020	.020	.189	.761	.381	.948	.101	.138	0.252	0.582
S-MRT-2	.129	.913	.033	.976	.860	.999	.009	.015	0.437	0.803
S-MRT-3	.336	.616	.045	.925	.623	.989	.027	.048	0.341	0.685

Anmerkung: Unterstreichungen kennzeichnen die besten Resultate je Kriterium

Tabelle 75 führt für die drei ausgewählten Verfahren die mittleren absoluten Abweichungen vom nominalen Niveau sowie die mittleren Intervalllängen getrennt nach Modell, Stichprobengröße und Parameterwert auf, wobei die je Bedingung günstigsten Werte unterstrichen sind. Insgesamt zeigt sich hier, daß mit steigendem Stichprobenumfang bei allen Verfahren die empirischen Überdeckungswahrscheinlichkeiten näher am nominalen Wert liegen und die Intervalle durchweg kürzer ausfallen. Unter den stetigen Verteilungen zeigt auch hier das Shift-Verfahren im Vergleich das gleichmäßigste Verhalten, der Überdeckungsfehler ist relativ unabhängig vom Parameterwert. Zudem fallen die Intervalle relativ kürzer aus als bei den anderen beiden Verfahren. Unter dem hypergeometrischen Modell ist der absolute Überdeckungsfehler des Shift-Verfahrens für Parameterwerte nahe .5 auch hier deutlich größer als unter stetigen Verteilungen. Für die beiden anderen Verfahren ergeben sich hinsichtlich der absoluten Abweichungen vom nominalen Niveau zwischen den Verteilungen kaum Unterschiede. Korrespondierend zu diesem Befund zeigt sich, daß die Intervalle des Shift-Verfahrens bei $\theta_b = .5$ und $\theta_b = .611$ unter dem diskreten Modell kürzer ausfallen, die Intervalle von CLI-PS-T und CLI-LOG-T hingegen länger. Die Logit-Transformation zeigt auf dem Aggregationsniveau von Tabelle 75 unter keiner Bedingung einen praktischen Vorteil gegenüber der Pseudo-Score-Formel.

Tabelle 75. Mittlere absolute Abweichung der simulierten von der nominalen Überdeckungswahrscheinlichkeit und Intervalllänge ausgewählter Verfahren

	n	θ_b	$ \overline{CP} - .95 $			Mittlere Intervalllänge		
			CLI-PS-T	CLI-LOG-T	S-CLI-T-3	CLI-PS-T	CLI-LOG-T	S-CLI-T-3
Stetig	≤ 12	.500	.020	.026	<u>.017</u>	<u>0.464</u>	0.490	0.465
		.611	<u>.020</u>	.025	.021	0.457	0.484	<u>0.454</u>
		.714	.024	.024	<u>.018</u>	0.436	0.469	<u>0.428</u>
		.802	.043	.033	<u>.016</u>	0.400	0.442	<u>0.395</u>
		.871	.085	.058	<u>.022</u>	<u>0.346</u>	0.397	0.359
	≥ 14	.500	.012	.017	<u>.011</u>	<u>0.334</u>	0.343	<u>0.334</u>
		.611	<u>.011</u>	.015	<u>.011</u>	0.328	0.336	<u>0.326</u>
		.714	<u>.009</u>	.014	.011	0.308	0.317	<u>0.307</u>
		.802	<u>.007</u>	.009	.011	0.279	0.291	<u>0.277</u>
		.871	.020	.012	<u>.010</u>	0.241	0.258	<u>0.240</u>
Hyper	≤ 12	.500	<u>.024</u>	.030	.032	0.506	0.540	<u>0.429</u>
		.611	<u>.020</u>	.027	.030	0.473	0.502	<u>0.437</u>
		.714	.024	<u>.022</u>	.026	0.434	0.466	<u>0.428</u>
		.802	.036	.026	<u>.016</u>	<u>0.391</u>	0.430	0.406
		.871	.082	.052	<u>.015</u>	<u>0.343</u>	0.391	0.389
	≥ 14	.500	<u>.015</u>	.021	.029	0.370	0.382	<u>0.319</u>
		.611	<u>.013</u>	.018	.018	0.340	0.353	<u>0.324</u>
		.714	<u>.009</u>	.014	.015	<u>0.307</u>	0.316	0.310
		.802	<u>.006</u>	.009	.012	<u>0.272</u>	0.282	0.285
		.871	.022	.010	<u>.007</u>	<u>0.240</u>	0.255	0.261

Anmerkung: Unterstreichungen kennzeichnen die besten Resultate je Kriterium

Hinsichtlich der Überdeckungswahrscheinlichkeit ermöglichen Abbildung 48 und Abbildung 49 einen detaillierten Vergleich. Zunächst wird deutlich, daß die Überdeckungskurven von CLI-PS-T und CLI-LOG-T sehr ähnliche Verläufe zeigen, wobei CLI-LOG-T in aller Regel etwas konservativer ist als CLI-PS-T.

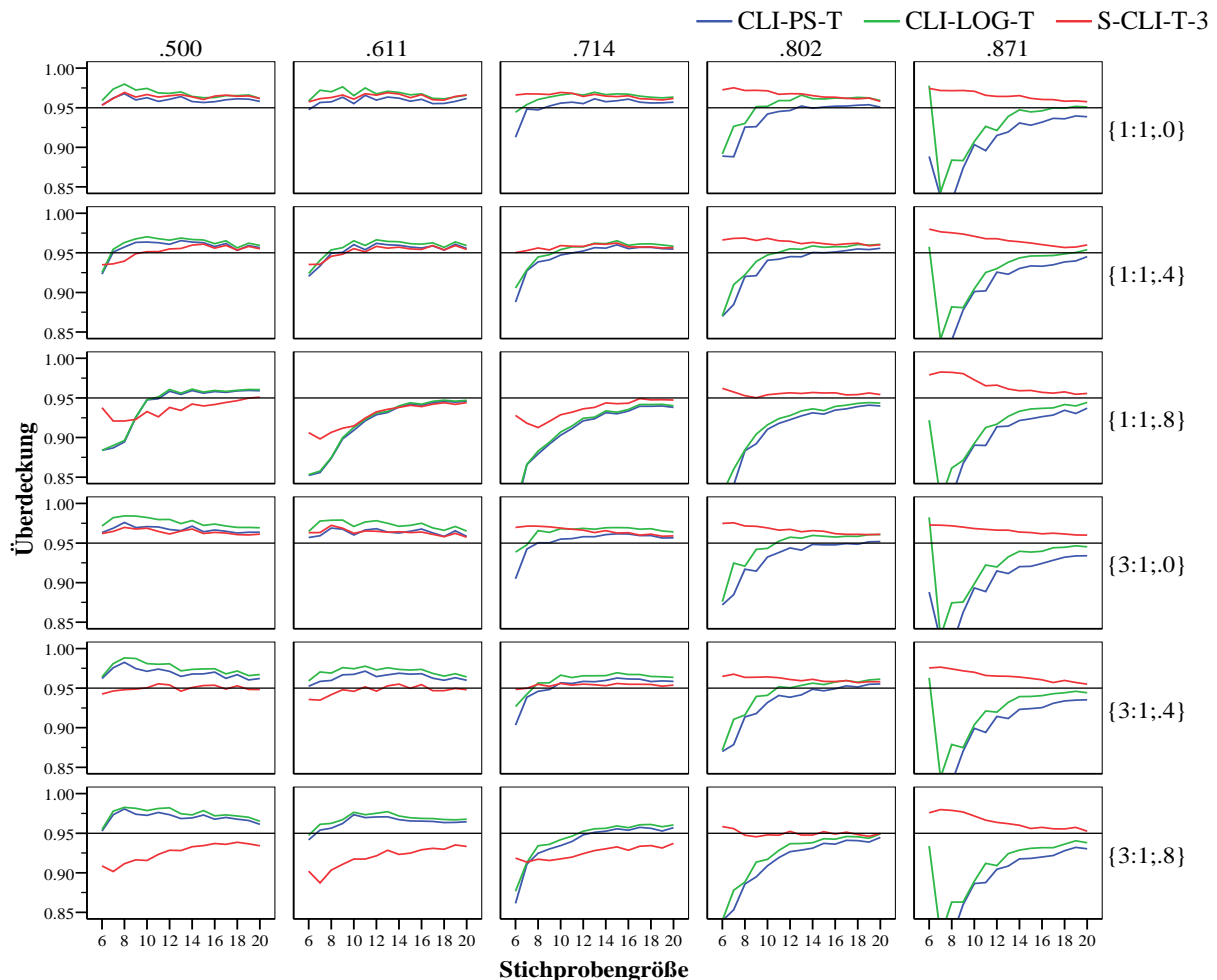


Abbildung 48. Mittlere Überdeckungswahrscheinlichkeiten von CLI-PS-T, CLI-LOG-T und S-CLI-T-3 unter stetigen Verteilungen

Das Shift-Intervall hingegen zeigt deutlich andere Verläufe. Während CLI-PS-T und CLI-LOG-T bei mittleren Parameterwerten ($\theta_b = .5$ und $\theta_b = .611$) sehr robust überdecken und erst bei großen Parameterwerten und insbesondere kleinen Stichproben das nominale Niveau unterschreiten, weist das Shift-Intervall bei diesen Parameterwerten und vor allem in Verbindung mit einer hohen Korrelation zwischen den Gruppen die größten Überdeckungsfehler auf. Auch im abhängigen Fall erweist sich das Shift-Verfahren unter den diskreten Modellen nur teilweise geeignet zur Schätzung der stochastischen Überlegenheit. Bei heterogenen und korrelierten Verteilungen ist dieses Verfahren auch bei $n = 20$ hierfür nicht geeignet, da die nominale Überdeckungswahrscheinlichkeit zum Teil deutlich unterschritten wird. Für alle anderen Bedingungen deutet sich an, daß die empirische Überdeckungswahrscheinlichkeit mit steigendem n gegen den nominalen Wert strebt. Das Überdeckungsverhalten von CLI-PS-T und CLI-LOG-T unterscheidet sich zwischen stetigen und hypergeometrischen Verteilungen hingegen kaum, wobei sich auch hier andeutet, daß die Schätzungen mit steigendem n das Niveau zunehmend gut einhalten.

Bis zu einem $n = 20$ kann aufgrund der vorliegenden Analyse für die stetigen Verteilungen also das Shift-Verfahren empfohlen werden, wobei nicht robustes Verhalten hier nur bei hoher Korreliertheit in Verbindung mit kleinen Effekten zu erwarten ist. Für diskrete Modelle kann das Verfahren aufgrund der vorliegenden Untersuchung nicht generell empfohlen werden, in einzelnen Fällen liegt die empirische Überdeckungswahrscheinlichkeit für $n = 20$ unter .90 bei einem nominalen Wert von .95. Für größere Stichproben deuten die Simulationen die Möglichkeit einer robusten Schätzung sowohl für stetige als auch für diskrete Modelle anhand der Verfahren nach Cliff an. Dies sollte in weiteren Simulationsstudien untersucht werden.

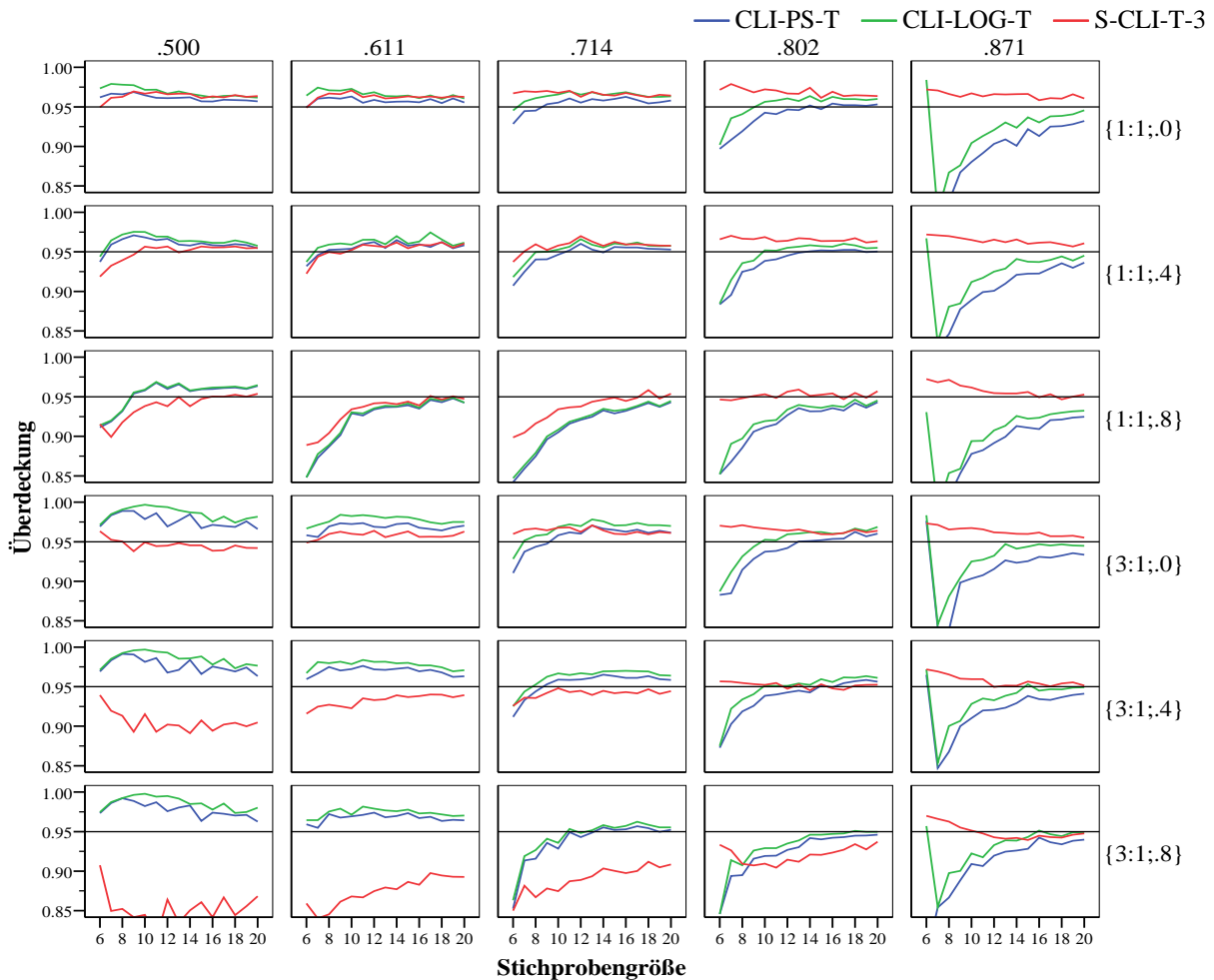


Abbildung 49. Überdeckungswahrscheinlichkeiten von CLI-PS-T, CLI-LOG-T und S-CLI-T-3 unter negativen hypergeometrischen Verteilungen

Die Boxplots der Überdeckungswahrscheinlichkeiten in Abbildung 50 unterstreichen, daß S-CLI-T-3 bei den hier betrachteten Stichprobengrößen insgesamt und auch getrennt nach Verteilungsmodellen näher am nominalen Niveau liegt und zudem keine längeren Intervalle aufweist als die beiden Verfahren mit dem Varianzschätzer nach Cliff. Mit den genannten empirischen aber auch theoretischen (vgl. Kapitel 3.2.2.2.3) Einschränkungen kann also für kleine Stichprobengrößen von den hier untersuchten Verfahren am ehesten das Shift-Intervall zur Schätzung der stochastischen Überlegenheit bei abhängigen Stichproben empfohlen werden.

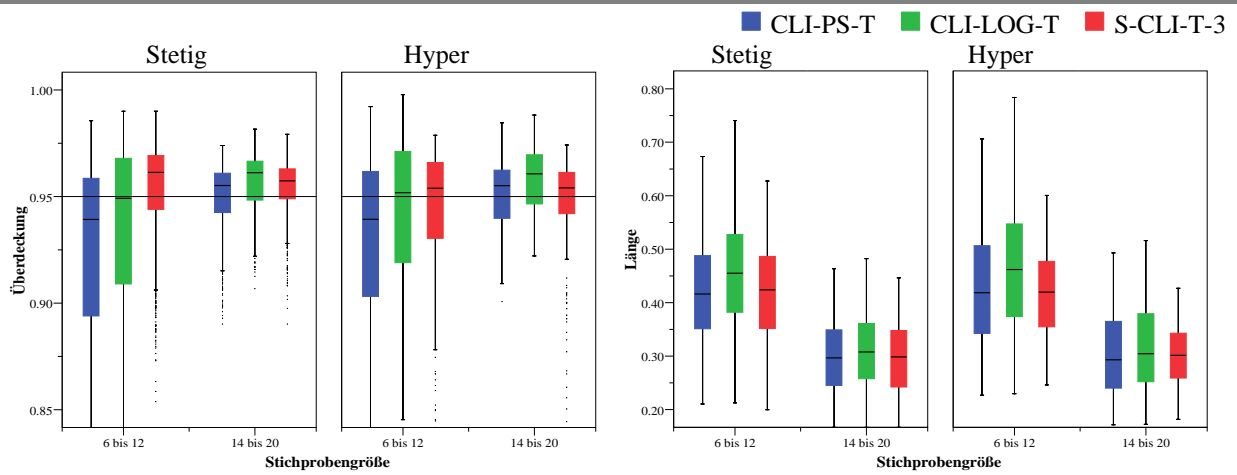


Abbildung 50. Simulierte Überdeckungswahrscheinlichkeiten und Intervallängen von CLI-PS-T, CLI-LOG-T und S-CLI-T-3

4 Anwendungen

Da bei den folgenden Anwendungsbeispielen keine multiplen Paarvergleiche gemacht werden und zudem keine expliziten Kontroll- oder Vergleichsgruppen vorliegen, wird auf die Demonstration der Effektgrößen nach Glass (1976), bei denen an nur einer Gruppenstreuung standardisiert wird, verzichtet.

4.1 Fahreignung und Spurhalten⁹

In dem Arbeitsprogramm-Projekt „Untersuchung der Fahreignung von Patienten mit hepatischer Enzephalopathie mit dem Fahrzeug für Interaktionsforschung/Straßenverkehr (FIS) und am Fahr Simulator“ der Bundesanstalt für Straßenwesen (BASt)¹⁰ wurden gesunde und leberkranke Fahrerlaubnisinhaber zwischen 31 und 81 Jahren medizinisch, computerpsychometrisch (CPM) sowie anhand von Realfahrtversuchen mit dem FIS untersucht. In den an den Universitätskliniken Düsseldorf durchgeführten psychometrischen Untersuchungen wurde unter anderem die verkehrspsychologische Testbatterie zur Leistungsuntersuchung nach der deutschen Fahrerlaubnisverordnung (FeV Anlage 5 Nr. 2) angewendet und zur (ex post facto) Einteilung der Pbn in eine Gruppe geeigneter ($n_1 = 79$) und eine Gruppe ungeeigneter ($n_2 = 20$) Fahrer genutzt. Diese Testbatterie besteht aus fünf Verfahren (Cognitrone, Determinationstest, Linienverfolgungstest, Reaktionstest, Tachistoskopischer Verkehrsauffassungstest) des Wiener Testsystems der Firma Schuhfried.

Die Realfahrtversuche mit dem FIS wurden auf dem ADAC-Verkehrsübungsplatz in Kaarst durchgeführt und umfaßten insgesamt 14 Fahraufgaben, in denen eine Vielzahl verkehrsrelevanter Situationen nachgestellt wurde. In der im folgenden betrachteten Aufgabe "Cassette suchen" mußten die Pbn mit vorgeschriebener Geschwindigkeit durch eine mit Pylonen abgesteckten Gasse fahren und nach Aufforderung durch den Versuchsleiter in der Mittelkonsole eine bestimmte Musikkassette finden und in das Abspielgerät einlegen. Eine wesentliche abhängige Variable dieses Versuches ist die Fähigkeit der Pbn, das Fahrzeug während der Suche in der Gassenspur zu halten. Innerhalb der Gasse wurden die Abstände des Fahrzeugs zum Gassenrand bestimmt. Anhand dieser Abstandsmessung läßt sich ein idealer (gerader) Kurs durch die Gasse bestimmen. Als Indikator des Spurhaltens wurde die Summe der absoluten Abweichungen (in cm) von diesem individuell berechneten Kurs gewertet. Tabelle 76 zeigt die Spurhalte-daten getrennt nach psychometrisch diagnostizierter Fahreignung.

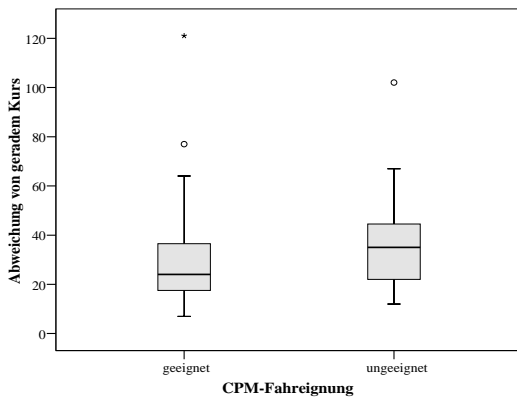
Tabelle 76. Spurhalte-daten der Aufgabe "Cassette suchen" (X_1 : geeignet; X_2 : ungeeignet)

X_1	7	9	9	10	11	12	12	12	13	15	15	15	15	16	16	16	16	17	17	17
	18	18	18	19	19	19	20	21	21	21	21	21	21	21	21	22	22	22	23	24
	24	25	26	26	26	28	28	29	29	29	30	31	31	31	31	32	33	35	36	37
	37	37	37	38	38	38	41	41	42	42	44	45	46	48	60	61	64	77	121	
X_2	12	14	15	19	21	23	32	33	33	35	35	39	40	43	44	45	49	52	67	102

⁹ Dieser Dissertation liegen Teile der im Auftrag des Bundesministeriums für Verkehr, Bau und Stadtentwicklung, vertreten durch die Bundesanstalt für Straßenwesen, unter FE-Nr. 82.0199/2002/AP durchgeführten Forschungsarbeit zugrunde. Die Verantwortung für den Inhalt liegt allein beim Autor.

¹⁰ Für die freundliche Überlassung der Daten sei insbesondere Frau Dr. Knoche von der BASt gedankt.

Die Boxplots und die Kennwerte in Abbildung 51 geben die Verteilungen der Spurhaltesdaten getrennt nach Fahreignung wieder, wobei deutlich wird, daß die als ungeeignet diagnostizierten Pbn tendentiell größere Abweichungen von einem geraden Kurs aufweisen. Im Sinne der Ausreißerdefinition nach Tukey (1977) hat die rechtsschiefe Verteilung der geeigneten Pbn zwei Ausreißer (bzw. einen Ausreißer und einen Extremwert), die ebenfalls rechtsschiefe Verteilung der ungeeigneten Pbn weist einen Ausreißer auf. Diese Werte sind dafür verantwortlich, daß die robustifizierten Lagedifferenzen sowie die stochastische Überlegenheit wie im folgenden dargestellt auf größere – und signifikante – Effekte hindeuten als die KQ-Statistiken.



	geeignet	ungeeignet
Anzahl	79	20
Mittelwert	28.304	37.650
Standardabweichung	17.112	20.681
Minimum	7.000	12.000
Median	24.000	35.000
Maximum	121.000	102.000
Schiefe	2.567	1.591
Kurtosis	10.623	4.046
20%-getrimmtes Mittel	25.270	35.250
20%-winsorisierte Streuung	8.389	9.783

Abbildung 51. Verteilungen der Spurhaltesdaten

Die nichtkorrigierten Punktschätzungen der Effektgrößen für den Fall heterogener Varianzen lauten $d'_s = -0.492$ und $d'_{s,R}{}^{Algina} = -0.703$, die anhand der t -Verteilung korrigierten Werte sind $\hat{d}' = -0.488$ und $\hat{d}'_R{}^{Algina} = -0.691$. Tabelle 77 gibt die Grenzen der nichtzentralen sowie der jeweils robustesten Bootstrap-Verfahren aus Studie 1 (vgl. Kapitel 2.1.2) an. Während die Intervalle der KQ-Statistiken die Null umschließen, liegen die robustifizierten Intervalle im negativen Bereich.

Tabelle 77. 95%-Konfidenzintervalle für d' und $d'_R{}^{Algina}$

Grenzen	NC(d'_s)	BCA(\hat{d}')	NC($d'_{s,R}{}^{Algina}$)	Perc($\hat{d}'_R{}^{Algina}$)
unten	-1.022	-0.965	-1.303	-1.460
oben	0.046	0.035	-0.084	-0.141

Die Mann-Whitney-Statistik ist hier $U = 529$, so daß die stochastische Überlegenheit $\hat{\theta} = .335$ ist. Tabelle 78 gibt die Intervallschätzungen der drei besten Verfahren aus Studie 3 (vgl. Kapitel 3.1.6) an. Alle zugehörigen ungerichteten Tests führen auf $\alpha = .05$ also zur Annahme der H_1 , so daß jeweils von stochastischer Ungleichheit der zugrundeliegenden Verteilungen auszugehen wäre.

Tabelle 78. 95%-Konfidenzintervalle für die stochastische Überlegenheit

Grenzen	MEE-PS-1	NEW-M-3	S-BM-Z-3
unten	.211	.210	.223
oben	.487	.478	.494

4.2 Progressive-Matrizentests mit und ohne Speed-Komponente

In den experimentellen Untersuchungen zu Invarianzhypothesen des Kurzzeitgedächtnisses von Klein (1995) wurde zur Überprüfung des Zusammenhangs zwischen psychometrischer Testintelligenz und Kurzzeitgedächtnis u.a. der Advanced-Progressive-Matrizentest (APM) von Raven als Indikator der fluiden Intelligenz appliziert. Zur Vermeidung eines Deckeneffektes wurde der APM zunächst als Speed-Test eingesetzt, wobei den insgesamt 38 Pbn für die Bearbeitung des ersten Matrizen-Sets zehn Minuten und für das zweite Set 40 Minuten Zeit eingeräumt wurde. Danach wurde der APM in herkömmlicher Weise als Power-Test durchgeführt, wobei den Pbn beliebig viel Zeit zur Be- oder auch Überarbeitung aller Items beider Sets zur Verfügung stand. Tabelle 79 gibt die Anzahl gelöster Matrizen unter der Speed- (X_1) - und der Power- (X_2) -Bedingung wieder¹¹.

Tabelle 79. Anzahl gelöster APM-Items mit (X_1) und ohne (X_2) Zeitbegrenzung aus Klein (1995)

<i>i</i>	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
X_1	29	30	30	30	31	32	32	33	33	35	35	35	36	36	36	36	38	38	38	39
X_2	29	35	30	31	40	34	33	33	36	35	36	39	40	38	46	36	41	38	42	43
<i>i</i>	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38		
X_1	39	39	39	39	40	40	40	41	41	42	42	43	43	43	44	45	45	45		
X_2	40	42	39	42	40	42	45	41	41	42	42	47	45	46	46	47	47	46		

Wie die Kennwerte in Tabelle 80 zeigen, sind die beiden Verteilungen relativ symmetrisch und weisen gemäß der Kurtosis kurze Tails auf; die getrimmten Mittelwerte weichen hier nur unwesentlich von den KQ-Schätzungen ab. Zudem weisen beide Schätzerarten auf homogene Streuungen und hohe Korrelation der beiden Stichproben hin.

Tabelle 80. Kennwerte der APM-Daten von Klein (1995)

	X_1	X_2
Anzahl	38	38
Mittelwert	37.684	39.868
Standardabweichung	4.674	4.987
Schiefte	-0.225	-0.423
Kurtosis	-0.928	-0.582
20%-getrimmtes Mittel	37.939	40.219
20%-winsorisierte Streuung	3.467	3.569
Korrelation	.882	
20%-winsor. Korrelation	.848	

Die Leistungssteigerung von durchschnittlich etwa zwei Items von der Speed- zur Power-Bedingung wird von Klein als geringfügig bezeichnet (S. 179). Wie beschreiben nun die verschiedenen Effektgrößenschätzer der vorliegenden Arbeit diese Leistungsverbesserung?

¹¹ Für die Überlassung der Daten sei Herrn Dr. Klein herzlichst gedankt.

Zunächst seien die Statistiken zum intraindividuellen Vergleich (d.h. unter Berücksichtigung der Korrelation) dargestellt. Die unkorrigierte Schätzung der KQ-Lagedifferenz ist $d_{D,s} = -0.922$ ($t_{\text{emp}} = -5.683$; $v = 37$); die robustifizierte Schätzung fällt mit $d_{D,s,R}^{\text{Algina}} = -0.755$ ($t_{\text{emp}} = -4.310$; $\hat{v} = 21.8$) etwas niedriger aus. Die aufgrund von Studie 2 (vgl. Kapitel 2.2.5.1) zu empfehlenden biaskorrigierten Punktschätzungen lauten $\hat{d}_D = -0.903$ (bzw. $\hat{d}_D^{\text{Boot}} = -0.869$) und $\hat{d}_{D,R}^{\text{Algina}} = -0.729$. Tabelle 81 führt für beide Parameter die nichtzentralen Intervallschätzungen sowie die jeweils (insgesamt) robusteren Bootstrap-Intervalle auf.

Tabelle 81. 95%-Konfidenzintervalle für d_D und $d_{D,R}^{\text{Algina}}$

Grenzen	NC ($d_{D,s}$)	BCA (\hat{d}_D)	NC ($d_{D,s,R}^{\text{Algina}}$)	Perc ($\hat{d}_{D,R}^{\text{Algina}}$)
unten	-1.298	-1.189	-1.156	-1.107
oben	-0.537	-0.657	-0.340	-0.524

Die Schätzung der stochastischen Überlegenheit zur intraindividuellen Veränderung ergibt hier $\hat{\theta}_w = .158$. Da insgesamt $N_0 = 12$ Bindungen auftreten, werden in Tabelle 82 die Schätzungen von Cliff-PS und Agresti(N_0) angegeben (vgl. Kapitel 3.2.3), wobei zum Vergleich auch das originale Agresti-Verfahren aufgeführt ist, bei welchem von bindungsfreien, stetigen Daten ausgegangen wird.

Tabelle 82. 95%-Konfidenzintervalle für θ_w

Grenzen	Cliff-PS	Agresti(N_0)	Agresti
unten	.097	.107	.071
oben	.246	.272	.308

Da die Korrelationen deutlich höher als .5 sind (vgl. Tabelle 80), weisen die Schätzer der interindividuellen Vergleiche auf deutlich kleinere Effekte hin: Die unkorrigierten und biaskorrigierten Lagedifferenzen lauten $d'_s = -0.452$, $\hat{d}'_{cl} = -0.440$, $d'_{s,R}{}^{\text{Algina}} = -0.416$ und $\hat{d}'_{cl,R}{}^{\text{Algina}} = -0.402$. Tabelle 83 führt die zugehörigen Intervallschätzungen auf.

Tabelle 83. 95%-Konfidenzintervalle für d' und $d'_{cl,R}{}^{\text{Algina}}$

Grenzen	NC ₁ (d'_s)	BCA (\hat{d}'_{cl})	NC ₁ ($d'_{s,R}{}^{\text{Algina}}$)	Perc ($\hat{d}'_{cl,R}{}^{\text{Algina}}$)
unten	-0.636	-0.646	-0.636	-0.697
oben	-0.263	-0.292	-0.186	-0.244

Die Schätzung der stochastischen Überlegenheit zwischen den Gruppen schließlich ist $\hat{\theta}_b = 0.375$, die Intervallschätzungen der beiden besten Verfahren sind in Tabelle 84 aufgeführt.

Tabelle 84. 95%-Konfidenzintervalle für θ_b

Grenzen	CLI-LOG-T	S-CLI-T-3
unten	.324	.317
oben	.429	.424

Sämtliche Verfahren führen bei diesen (ausreißerfreien) Daten also zu sehr ähnlichen Ergebnissen, wobei die Schätzungen unter Berücksichtigung der Korrelation auf einen *großen Effekt* (gemäß der Konvention nach Cohen) hindeuten und die Schätzungen, bei denen die Korrelation ignoriert bzw. herausgerechnet wird, zeigen einen *mittleren Effekt* an. Alle Intervallschätzungen korrespondieren zu signifikanten Testergebnissen (jeweils zweiseitig mit $\alpha = .05$).

5 Zusammenfassung

Vor dem Hintergrund, daß der Schätzung von Effektgrößen in der psychologischen Forschung ein zunehmender Stellenwert zukommt, wurden in der vorliegenden Arbeit für zwei grundlegende Datensituationen, dem Zwei-Gruppen-Fall mit unverbundenen und dem Zwei-Gruppen-Fall mit verbundenen Stichproben, die Eigenschaften von Punkt- und insbesondere Intervallschätzern für standardisierte Lagedifferenzen und die stochastische Überlegenheit anhand von Monte-Carlo-Simulationen untersucht. Das wesentliche Ziel hierbei war es, geeignete Schätzer für das (verallgemeinerte) Lagen-Skalen-Modell zu finden, also Verfahren, die sich auch unter nichtnormalen Verteilungen sowie bei heterogenen Varianzen (relativ) robust verhalten.

In den Studien zu standardisierten Lagedifferenzen wurden sowohl nichtzentrale als auch Bootstrap-Schätzer (Perzentil-, BCA- und Perzentil- t -Bootstrap) für KQ- und 20%-robustifizierte Effektgrößen untersucht. Da sich die Ergebnisstrukturen der Studien zu unverbundenen und verbundenen Stichproben weitgehend gleichen, seien diese getrennt nach Punkt- und Intervallschätzung wie folgt zusammengefaßt:

- Die unkorrigierten KQ-Punktschätzer führen unter Verteilungen mit stärkeren Tails mit steigendem Parameterwert zu zunehmender Überschätzung (positiver Bias) der Parameter. Diese Verzerrungen treten auch bei balancierten Gruppengrößen auf und können nicht durch die Biaskorrekturen auf Basis der t -Verteilung (Huynh, 1989) behoben werden. Die Bootstrapbiaskorrigierten KQ-Schätzer weisen insgesamt geringere Verzerrungen auf. Bei Disbalance – das Folgende betrifft also nur den unabhängigen Fall – führen die Bootstrap-Schätzer unter Verteilungen mit kurzen oder normalen Tails jedoch zu teilweise deutlichen Unterschätzungen. Mit den hier untersuchten Statistiken stehen somit keine generell empfehlenswerten KQ-Schätzer zur Verfügung.
- Für die robustifizierten Punktschätzer fallen die Verzerrungen unter den tailstärkeren Verteilungen durchweg geringer aus als für die KQ-Statistiken. Für die Parameter d'_R und $d_{D,R}$ weisen die t -basierten Korrekturen und Bootstrap-Korrekturen ähnliche Verzerrungsquadrate (B_{MC}^2) auf, wobei die t -korrigierten Statistiken eher positive und die Bootstrap-Korrekturen eher negative Verzerrungen zeigen. Anders als bei den KQ-Statistiken fallen die mittleren quadratischen Fehler (MSE_{MC}) der t -korrigierten Schätzer geringer aus als diejenigen der Bootstrap-Schätzer. Während diese Beschreibung bei verbundenen Stichproben auch auf die jeweiligen Schätzer für $d_R^{S_k}$ zutrifft, zeigt sich bei unverbundenen Stichproben, daß das Bootstrap als Biaskorrektur nicht geeignet ist, wenn die Varianzschätzung auf nur zehn Werten beruht. In diesem Fall kann es zu extremen Verzerrungen kommen, so daß bei kleinen Stichproben generell die t -Korrektur vorzuziehen ist.
- Auch die Ergebnisse zu den Intervallschätzern zeigen für die verschiedenen Parameter ein recht einheitliches Muster. Sowohl für die KQ- als auch für die robustifizierten Effektgrößen weisen jeweils einzelne Bootstrap-Verfahren robusteres Überdeckungsverhalten auf als die nichtzentralen Intervallschätzer. Während die Schätzung der KQ-Effektgrößen für alle betrachteten Parameter insgesamt, also über alle Verteilungen hinweg, mit dem BCA-Bootstrap am besten gelingt, ist für die Schätzung der robustifizierten Parameter das Perzentil-Bootstrap zu empfehlen.
- Wie die Gegenüberstellungen der KQ- und der robustifizierten Schätzungen deutlich machen (siehe S. 66 f. und S. 102 f.), überdecken die Schätzer für KQ- und für robustifizierte Effektgrößen unter den Verteilungen mit kurzen oder normalen Tails ähnlich gut. Hingegen kann mit den hier betrachteten Verfahren bereits für eine Verteilung ($g-h-2$) mit der Charakteristik der Exponentialverteilung keine robuste Schätzung von KQ-Effektgrößen vorgenommen werden. Die

Schätzung der robustifizierten Größen gelingt auch unter den tailstarken Verteilungen recht robust. Nachteilig bei der Schätzung der robustifizierten Effektgrößen ist jedoch, daß die Intervalle unter Verteilungen mit normalen oder kurzen Tails deutlich breiter ausfallen als die KQ-Schätzungen (beachte in diesem Zusammenhang Kapitel 2.1.5.1 und 2.1.5.2 zur einheitlichen Skalierung der Schätzungen verschiedener Parameter).

- Die Empfehlungen für die Intervallschätzung standardisierter Lagedifferenzen (siehe Kapitel 2.1.5.3 und 2.2.5.3) werden daher von den Annahmen über die Stärke der Verteilungstails abhängig gemacht: Für Verteilungen mit normalen oder kurzen Tails bieten sich die nichtzentralen Schätzer an und für Situationen, in denen auch von Verteilungen mit stärkeren Tails auszugehen ist, wird eine robustifizierte Intervallschätzung anhand des Perzentil-Bootstraps empfohlen.

Hinsichtlich der simulierten Überdeckungswahrscheinlichkeiten stehen die vorgelegten Ergebnisse weitgehend im Einklang mit der einschlägigen Literatur (vgl. Kapitel 2.1.3 und 2.2.3). Der generellen Empfehlung zur (20%-)robustifizierten Schätzung (vgl. Algina, Keselman & Penfield, 2005a, 2005b) wird hier aufgrund der mangelhaften Schätzpräzision unter Verteilungen mit kürzeren Tails jedoch nicht gefolgt. Ein erster Lösungsansatz für dieses Problem wäre die Verringerung des Trimmanteils. Anhand von Simulationen wäre ein Trimmanteil zu finden, der zu Intervallschätzungen führt, die einen geeigneten Kompromiß zwischen Robustheit und Präzision darstellen, wobei sich wiederum die Frage stellt, welches Spektrum von in der Psychologie relevanten Verteilungen zu berücksichtigen wäre (vgl. Kapitel 2.1.4.1). Ein zweiter Lösungsansatz könnte in der Entwicklung eines adaptiven Schätzers liegen. Beim adaptiven Schätzen und Testen (Büning, 1991; Büning & Rietz, 2003; Keselman, Wilcox, Lix, Algina & Fradette, 2007) wird in Abhängigkeit von der Verteilungscharakteristik der beobachteten Daten die Wahl eines optimalen Verfahrens angestrebt.¹² Bei der Verwendung getrimmter Mittelwerte könnte anhand einer Selektorstatistik – z.B. Tailmaße nach Hogg (1974) – der erforderliche Trimmanteil gewählt werden, der eine möglichst präzise Schätzung verspricht. Für die adaptive Schätzung mit variablem Trimmanteil ließe sich die Effektgröße nach Algina, Keselman und Penfield (2005a) so gestalten, daß unter der Normalverteilung die herkömmliche Effektgröße d' geschätzt wird, indem mit der jeweiligen γn %-winsorisierten Streuung der Normalverteilung multipliziert wird: $\hat{d}'_{\gamma R} = \sigma_{\gamma w}^{Normal} (\bar{X}_{\gamma r1} - \bar{X}_{\gamma r2}) / \sqrt{(S_{\gamma w1}^2 + S_{\gamma w2}^2) / 2}$. Problematisch bei einem solchen Ansatz wäre jedoch, daß in der praktischen Anwendung der tatsächliche Trimmanteil des geschätzten Parameters nicht bekannt wäre.

Als alternatives Maß der Effektgröße wurde in den Studien 3 und 4 die stochastische Überlegenheit betrachtet. Diese ordinale Statistik bietet zwei wesentliche Vorteile gegenüber den KQ-Statistiken, denn sie ist von vergleichbarer Robustheit wie die robustifizierten Statistiken und kann zudem erwartungstreu geschätzt werden. Die Ergebnisse zur Intervallschätzung lassen sich wie folgt zusammenfassen:

- Bei unabhängigen Stichproben kann die stochastische Überlegenheit im stetigen Fall bereits bei sehr kleinen Stichproben mit dem in dieser Arbeit entwickelten Shift-Intervall S-BM-Z-3 in aller Regel robust geschätzt werden ($pCP_{.936}^{1.000} = .943$). Zudem weist dieses Shift-Verfahren keine längeren Intervalle als die konkurrierenden, etwas weniger robusten Verfahren MEE-PS-1 und NEW-M-3 auf. Im diskreten (negativen hypergeometrischen) Fall gehört S-BM-Z-3 zwar wei-

¹² Die prinzipielle Vorgehensweise eines adaptiven Tests entspricht der früher verbreiteten und mitunter bis heute in Lehrbüchern (Rasch, Friese, Hofmann & Naumann, 2006) empfohlenen Praxis, die Anwendung eines parametrischen Tests von dem Ergebnis eines vorgeschalteten Anpassungstests abhängig zu machen (z.B. Anwendung des Welch-Tests nur bei signifikantem Ergebnis des Levene-Tests auf Varianzhomogenität). Solche zweistufigen Verfahren sind jedoch äußerst bedenklich, da das Gesamtniveau dieser bedingten Testung nicht kontrolliert wird (Büning, 1991).

terhin zu den besten Verfahren, jedoch wird das nominale Konfidenzniveau bei heterogenen Varianzen, die hier mit einer hohen Wahrscheinlichkeit von gebundenen Werten einhergehen, zum Teil sehr deutlich unterschritten. Daher wird für den diskreten Fall das Verfahren nach Mee (1990) empfohlen sowie auf das in dieser Arbeit nicht berücksichtigte Score-Intervall nach Ryu und Agresti (2008) verwiesen.

- Bei verbundenen Stichproben konnten für die Schätzung der (intraindividuellen) stochastischen Überlegenheit θ_w im stetigen Fall die sehr guten Eigenschaften des adjustierten Wald-Intervalls nach Agresti und Coull (1998) repliziert werden. Zudem wurde eine neue Variante [Agresti(N_0)] für den diskreten Fall vorgeschlagen, bei welcher die Varianzschätzung um die Anzahl der Bindungen korrigiert wird. Dieses Verfahren weist unter den negativen hypergeometrischen Verteilungen mit $|\overline{CP} - .95| = .013$ die geringsten Abweichungen vom nominalen Niveau auf.
- Die Schätzung der (interindividuellen) Effektgröße θ_b gelingt weniger gut als bei unverbundenen Stichproben. Zwar kann auch hier eines der Shift-Intervalle (S-CLI-T-3) insgesamt, d.h. sowohl bei stetigen ($pCP_{.936}^{1.000} = .865$) als auch bei negativen hypergeometrischen Verteilungen ($pCP_{.936}^{1.000} = .778$) im Vergleich zu den anderen Schätzern als robustestes Verfahren bezeichnet werden, jedoch treten hier auch unter stetigen Verteilungen und insbesondere bei hoher Korrelation größere Unterschreitungen des nominalen Niveaus auf. Da die relativen Vorteile von S-CLI-T-3 auf das Verhalten bei sehr kleinen Stichproben zurückgehen, wird für Stichproben ab $n = 20$ das Verfahren CLI-PS-T nach Cliff (1996a) empfohlen.

Abschließend seien die simulierten Überdeckungswahrscheinlichkeiten der relativ besten Verfahren aller vier Studien für den Fall mit $N = 40$ bzw. $n = 20$ gegenübergestellt. Die Lagedifferenzen vom Glass'schen Typ werden hier nicht berücksichtigt (siehe hierzu die Kapitel 2.1.5.2.5 und 2.2.5.2.5). In Abbildung 52 und Abbildung 53 wird zusammenfassend deutlich, daß sowohl mit robustifizierten als auch mit ordinalen Verfahren über das gesamte Spektrum der simulierten Verteilungseigenschaften hinweg eine weitgehend robuste Schätzung der Effektgrößen im unverbundenen wie auch im verbundenen Fall möglich ist.

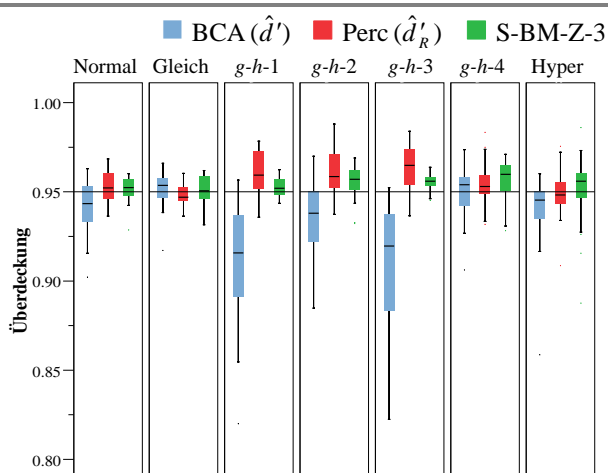


Abbildung 52. Gegenüberstellung der beobachteten Überdeckungswahrscheinlichkeiten ausgewählter Schätzer für die standardisierten Lagedifferenzen d' und d'_R sowie die stochastische Überlegenheit θ bei unverbundenen Stichproben mit $N = 40$

Zu beachten ist hierbei, daß die Shifts der Verteilungen nur unter dem Normalverteilungsmodell bei allen Studien identisch sind, da in den Studien 1 und 2 die Effektgröße d' und in den Studien 3 und 4 die stochastische Überlegenheit θ manipuliert wurde. Daher lassen sich die mit den Boxplots gezeigten Streuungen der simulierten Überdeckungswahrscheinlichkeiten zwischen den Verfahrensklassen auch nur unter dem Normalverteilungsmodell direkt miteinander vergleichen. Um einen solchen Vergleich der Überdeckungswahrscheinlichkeiten von robustifizierten und ordinalen Statistiken herbeizuführen, wären also Simulationen erforderlich, in denen alle Verfahren unter denselben Pseudozufallszahlen berechnet werden. In diesem Fall wäre dann auch ein Vergleich im Hinblick auf die Power der korrespondierenden Tests möglich, der als Ersatz zum Vergleich der Intervalllängen eingesetzt werden kann, wobei jeweils vorauszusetzen ist, daß die unter Betracht stehenden Verfahren sämtlich α -robust sind.

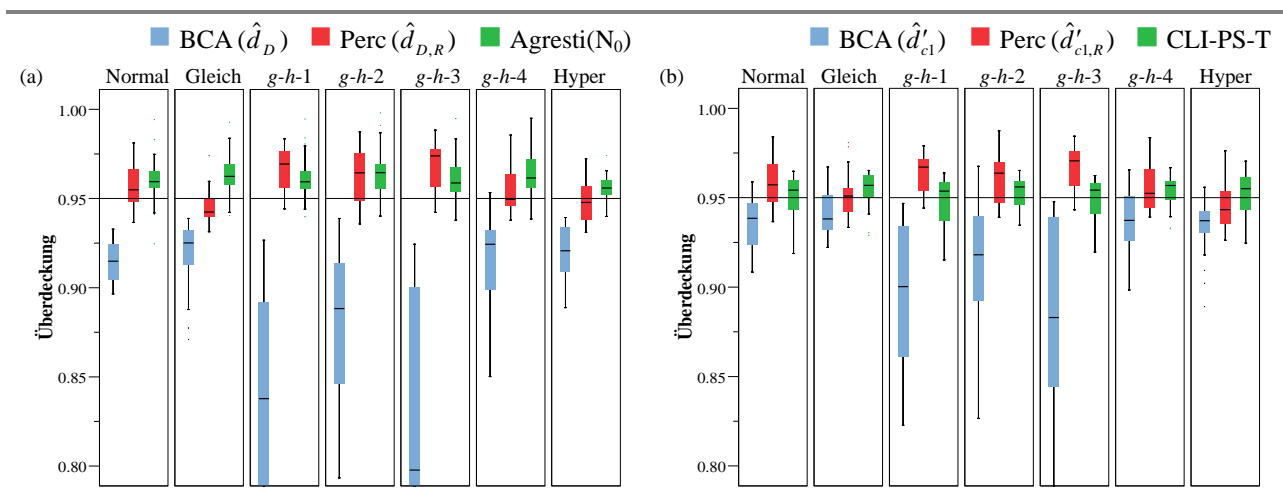


Abbildung 53. Gegenüberstellung der beobachteten Überdeckungswahrscheinlichkeiten ausgewählter Schätzer für die standardisierten Lagedifferenzen (a) d_D und $d_{D,R}$, (b) d' und $d'_{cl,R}$ sowie die stochastischen Überlegenheiten (a) θ_w und (b) θ_b bei verbundenen Stichproben mit $n = 20$

Die vorgelegten Untersuchungsergebnisse unterstreichen für die Schätzung von Effektgrößen insgesamt folgende Botschaft der Vertreter der ordinalen und robustifizierten Statistik (Büning, 1991; Cliff, 1996a; Wilcox, 2005): *Vermeide die Anwendung von KQ-Statistiken, wenn die Verteilungstails stärker als bei der Normalverteilung sind!* Obwohl aus statistischer Perspektive seit 40 Jahren gute Argumente für eine stärkere Beachtung robuster und (neuerer) rangbasierter Statistiken vorliegen, und die möglichen Konsequenzen nichtrobuster Verfahren – nämlich mangelnde Präzision und Power der Auswertung – den Forscher und den Forschungsprogreß konkret betreffen (Wilcox, 1998b), stehen der Verwendung der dargestellten Größen in der Forschungspraxis einige Hürden entgegen:

- Die Argumente gegen die KQ-Statistiken werden offensichtlich, nämlich seit geraumer Zeit, von der Wissenschaftsgemeinde nicht geteilt oder nicht als praxisrelevant erachtet;
- die herkömmliche Statistik ist elegant und hat sich (vermeintlich?) bewährt;
- neuere Verfahren der robusten Statistik finden nur sehr zögerlich Eingang in die bekannteren Statistikprogramme;
- die hier vorgestellten Schätzer sind vielfach nicht ohne Kenntnis der Primärdaten berechenbar, wodurch die Vergleichbarkeit mit anderen Studien eingeschränkt wird;

- zudem lassen sich diese Schätzer nicht ohne weiteres mit den bestehenden metaanalytischen Verfahren verknüpfen.

Mit diesen Hindernissen sind zugleich die Herausforderungen für die weitere Forschung genannt, die neben der Fortentwicklung der Verfahren insbesondere in dem an echten Daten zu führenden Nachweis der praktischen Relevanz robustifizierter und ordinaler Statistik liegen. Doch obwohl die Speicherung und elektronische Weitergabe von Daten heute kaum Schwierigkeiten bereiten sollten und die ethischen Prinzipien der wissenschaftlichen Gesellschaften die Weitergabe für Reanalysen auch ausdrücklich vorsehen, sind solche Reanalysen anscheinend heutzutage nicht leichter zu realisieren als früher. So konnte z.B. das eigentliche Ziel der Studie von Wicherts, Borsboom, Kats und Molenaar (2006), nämlich die Analyse der Verteilungseigenschaften von Daten aus in APA-Journals veröffentlichten Studien, aufgrund einer Nonresponserate von 73% seitens der kontaktierten Autoren überhaupt nicht erreicht werden. Der Forderung von Wicherts et al. nach einer obligatorischen Hinterlegung oder Veröffentlichung der Originaldaten als Voraussetzung seitens der Herausgeber für die Veröffentlichung empirischer Arbeiten ist in diesem Zusammenhang beizupflichten.

Literatur

- Agresti, A. & Caffo, B. (2000). Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. *The American Statistician*, 54, 280-288.
- Agresti, A. & Coull, B. A. (1998). Approximate is better than exact for interval estimation of binomial proportions. *The American Statistician*, 52, 119-126.
- Algina, J. & Keselman, H. J. (2003). Approximate confidence intervals for effect sizes. *Educational and Psychological Measurement*, 63, 537-553.
- Algina, J., Keselman, H. J. & Penfield, R. D. (2005a). An alternative to Cohen's standardized mean difference effect size: A robust parameter and confidence interval in the two independent groups case. *Psychological Methods*, 10, 317-328.
- Algina, J., Keselman, H. J. & Penfield, R. D. (2005b). Effect sizes and their intervals: The two-level repeated measures case. *Educational and Psychological Measurement*, 65, 241-258.
- Algina, J., Keselman, H. J. & Penfield, R. D. (2006a). Confidence intervals for an effect size when variances are not equal. *Journal of Modern Applied Statistical Methods*, 5, 2-13.
- Algina, J., Keselman, H. J. & Penfield, R. D. (2006b). Confidence interval coverage for Cohen's effect size statistic. *Educational and Psychological Measurement*, 66, 945-960.
- Ali, M. M. & Woo, J. (2005). Inference on $P(Y < X)$ in a pareto distribution. *Journal of Modern Applied Statistical Methods*, 4, 583-586.
- Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H. & Tukey, J. W. (1972). *Robust estimation of location: Survey and advances*. Princeton: University Press.
- Bauer, D. F. (1972). Constructing confidence sets using rank statistics. *Journal of the American Statistical Association*, 67, 687-690.
- Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology*, 41, 257-278.
- Beelmann, A. & Bliesener, T. (1994). Aktuelle Probleme und Strategien der Metaanalyse. *Psychologische Rundschau*, 45, 211-233.
- Berry, G. & Armitage, P. (1995). Mid-P confidence intervals: A brief review. *The Statistician*, 44, 417-423.
- Bird, K. (2002). Confidence intervals for effect sizes in analysis of variances. *Educational and Psychological Measurement*, 62, 197-226.
- Birnbaum, Z. W. (1956). On a use of the Mann-Whitney statistic. In J. Neyman (Ed.), *Proceedings of the third Berkeley symposium on mathematical statistics and probability. Vol. 1. Contributions to the theory of statistics* (pp. 13-17). Berkeley: University of California Press.
- Birnbaum, Z. W. & McCarty, R. C. (1958). A distribution-free upper confidence bound for $\Pr(Y < X)$, based on independent samples of X and Y. *Annals of Mathematical Statistics*, 29, 558-562.
- Blyth, C. R. (1986). Approximate binomial confidence limits. *Journal of the American Statistical Association*, 81, 843-855.
- Blyth, C. R. & Still, H. A. (1983). Binomial confidence intervals. *Journal of the American Statistical Association*, 78, 108-116.

- Box, G. E. P. & Muller, M. E. (1958). A note on the generation of random normal deviates. *Annals of Mathematical Statistics*, 29, 610-611.
- Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, 31, 144-152.
- Bredenkamp, J. (1969). Über Maße der praktischen Signifikanz. *Zeitschrift für Psychologie*, 177, 310-318.
- Bredenkamp, J. (1972). *Der Signifikanztest in der psychologischen Forschung*. Frankfurt a. M.: Akademische Verlagsanstalt.
- Bredenkamp, J. (1980). *Theorie und Planung psychologischer Experimente*. Darmstadt: Steinkopf.
- Bredenkamp, J. & Feger, H. (1970). Kriterien für die Entscheidung über die Aufnahme empirischer Arbeiten in die Zeitschrift für Sozialpsychologie. *Zeitschrift für Sozialpsychologie*, 1, 43-47.
- Bross, I. D. J. (1958). How to use riddit analysis. *Biometrics*, 14, 18-38.
- Brunner, E. & Munzel, U. (2000). The nonparametric Behrens-Fisher problem: Asymptotic theory and a small-sample approximation. *Biometrical Journal*, 42, 17-25.
- Brunner, E. & Munzel, U. (2002). *Nicht-parametrische Datenanalyse*. Berlin: Springer.
- Buchner, A., Erdfelder, E. & Faul, F. (1996). Teststärkeanalysen. In E. Erdfelder, R. Mausfeld, T. Meiser & G. Rudinger (Hrsg.), *Handbuch Quantitative Methoden* (S. 123-136). Weinheim: Psychologie Verlags Union.
- Bunke, O. (1959). Neue Konfidenzintervalle für den Parameter der Binomialverteilung. *Wissenschaftliche Zeitschrift der Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Reihe*, 9, 335-363.
- Bünig, H. (1991). *Robuste und adaptive Tests*. Berlin: Walter de Gruyter.
- Bünig, H. & Rietz, M. (2003). Adaptive bootstrap tests and their competitors in the c-sample location problem. *Journal of Statistical Computation and Simulation*, 73, 361-375.
- Capraro, R. M. & Capraro, M. M. (2002). Treatments of effect sizes and statistical significance tests in textbooks. *Educational and Psychological Measurement*, 62, 771-782.
- Chow, S. L. (1988). Significance test or effect size? *Psychological Bulletin*, 103, 105-110.
- Cliff, N. (1993). Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological Bulletin*, 114, 494-509.
- Cliff, N. (1996a). *Ordinal methods for behavioral data analysis*. Mahwah: Lawrence Erlbaum Associates.
- Cliff, N. (1996b). Answering ordinal questions with ordinal data using ordinal statistics. *Multivariate Behavioral Research*, 31, 331-350.
- Clopper, C. J. & Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26, 404-413.
- Cochran, W. G. (1937). Problems arising in the analysis of a series of similar experiments. *Journal of the Royal Statistical Society, Supplement*, 4, 102-118.
- Cochran, W. G. (1952). The χ^2 test of goodness of fit. *Annals of Mathematical Statistics*, 23, 315-345.
- Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology*, 65, 145-153.

- Cohen, J. (1965). Some statistical issues in psychological research. In B. B. Wolman (Ed.), *Handbook of clinical psychology* (pp. 95-121). New York: McGraw-Hill.
- Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale: Lawrence Erlbaum Associates.
- Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, *45*, 1304-1312.
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, *49*, 997-1003.
- Collani, E. v. & Dräger, K. (2001). *Binomial distribution handbook for scientists and engineers*. Boston: Birkhäuser.
- Cumming, G. & Finch, S. (2001). A primer on the understanding, use, and calculation of confidence intervals that are based on noncentral distributions. *Educational and Psychological Measurement*, *61*, 532-574.
- Cumming, G. & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist*, *60*, 170-180.
- Delaney, H. D. & Vargha, A. (2002). Computing several robust tests of stochastic equality with ordinaly scaled variables and small to moderate sized samples. *Psychological Methods*, *7*, 485-503.
- Deuchler, G. (1914). Über die Methoden der Korrelationsrechnung in der Pädagogik und Psychologie. *Zeitschrift für pädagogische Psychologie und experimentelle Pädagogik*, *15*, 141-131, 145-159, 229-242.
- Downton, F. (1973). The estimation of $\Pr(X < Y)$ in the normal case. *Technometrics*, *15*, 551-558.
- Edgington, E. S. (1966). Statistical inference and non-random samples. *Psychological Bulletin*, *66*, 485-487.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, *7*, 1-26.
- Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association*, *82*, 171-200.
- Efron, B. & Tibshirani, R. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.
- Erdfelder, E. (2004). Angewandte deduktivistische Methodologie in der empirischen Psychologie. In E. Erdfelder & J. Funke (Hrsg.), *Allgemeine Psychologie und deduktivistische Methodologie* (S. 98-123). Göttingen: Vandenhoeck & Ruprecht.
- Faul, F., Erdfelder, E., Lang, A. G. & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.
- Feng, D. & Cliff, N. (2004). Monte Carlo evaluation of ordinal d with improved confidence interval. *Journal of Modern Applied Statistical Methods*, *3*, 322-332.
- Fisher, R. A. (1925). *Statistical methods for research workers*. Edinburgh: Oliver & Boyd.
- Fisher, R. A. (1935). *The design of experiments*. Edinburgh: Oliver & Boyd.
- Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika*, *43*, 521-532.
- Fricke, R. & Treinies, G. (1985). *Einführung in die Meta-Analyse*. Bern: Huber.

- Gigerenzer, G. (1993). The superego, the ego, and the id in statistical reasoning. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues* (pp. 311-339). Hillsdale: Lawrence Erlbaum Associates.
- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5, 3-8.
- Glass, G. V., McGaw, B. & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills: Sage.
- Gleason, J. R. (1993). Understanding elongation: The scale contaminated normal family. *Journal of the American Statistical Association*, 88, 327-337.
- Govindarajulu, Z. (1968). Distribution-free confidence bounds for $P(X < Y)$. *Annals of the Institute of Statistical Mathematics*, 20, 229-238.
- Grissom, R. J. (1994). Probability of the superior outcome of one treatment over another. *Journal of Applied Psychology*, 79, 314-316.
- Grissom, R. J. & Kim, J. J. (2001). Review of assumptions and problems in the appropriate conceptualization of effect size. *Psychological Methods*, 6, 135-146.
- Grissom, R. J. & Kim, J. J. (2005). *Effect sizes for research*. Mahwah: Lawrence Erlbaum Associates.
- Haagen, K. & Seifert, H. G. (1979). *Methoden der Statistik für Psychologen (Band II)*. Stuttgart: Kohlhammer.
- Hager, W. (1992). *Jenseits von Experiment und Quasi-Experiment*. Göttingen: Hogrefe.
- Hager, W. (1996). Zusammenhangs- und Unterschiedsmaße. In E. Erdfelder, R. Mausfeld, T. Meiser & G. Rudinger (Hrsg.), *Handbuch Quantitative Methoden* (S. 87-98). Weinheim: Psychologie Verlags Union.
- Hager, W. (2000). Zur Wirksamkeit von Interventionsprogrammen: Allgemeine Kriterien der Wirksamkeit von Programmen in einzelnen Untersuchungen. In W. Hager, J.-L. Patry & H. Brezing (Hrsg.), *Evaluation psychologischer Interventionsmaßnahmen* (S. 153-168). Bern: Huber.
- Hager, W. (2004). *Testplanung zur statistischen Prüfung psychologischer Hypothesen*. Göttingen: Hogrefe.
- Hager, W. (2005). Vorgehensweisen in der deutschsprachigen psychologischen Forschung: Eine Analyse empirischer Arbeiten der Jahre 2001 und 2002. *Psychologische Rundschau*, 56, 191-200.
- Hager, W. & Westermann, R. (1983). Die Elle: 10 Jahre danach. *Zeitschrift für Sozialpsychologie*, 13, 250-252.
- Haller, H. & Krauss, S. (2002). Misinterpretations of significance: A problem students share with their teachers? *Methods of Psychological Research Online*, 7, 1-20.
- Halperin, M., Gilbert, P. R. & Lachin, J. M. (1987). Distribution-free confidence intervals for $\Pr(X_1 < X_2)$. *Biometrics*, 43, 71-80.
- Hanley, J. A. & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143, 29-36.

- Harlow, L., Mulaik, S. A. & Steiger, J. H. (1997). *What if there were no significance tests?* Mahwah: Lawrence Erlbaum Associates.
- Hartung, J. & Knapp, G. (2003). An alternative test procedure for meta-analysis. In R. Schulze, H. Holling & D. Böhning (Eds.), *Meta-analysis: New developments and applications in medical and social sciences* (pp. 53-69). Göttingen: Hogrefe & Huber.
- Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6, 107-128.
- Hedges, L. V. & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando: Academic Press.
- Hess, M. R. & Kromrey, J. D. (2002). Confidence intervals for the standardized mean difference: An empirical comparison of methods for interval estimation of effect sizes. *Paper presented at the American Educational Research Association annual conference*, New Orleans.
- Hess, M. R. & Kromrey, J. D. (2003). Confidence intervals for standardized mean differences: An empirical comparison of bootstrap methods under non-normality and heterogeneous variances. *Paper presented at the American Educational Research Association annual conference*, Chicago.
- Hess, M. & Kromrey, J. D. (2004). Robust confidence intervals for effect sizes: A comparative study of Cohen's *d* and Cliff's *delta* under non-normality and heterogeneous variances. *Paper presented at the American Educational Research Association annual conference*, San Diego.
- Hoaglin, D. C. (1985). Summarizing shape numerically: The g-and-h distributions. In D. C. Hoaglin, F. Mosteller & J. W. Tukey (Eds.), *Exploring data tables, trends, and shapes* (pp. 461-513). New York: Wiley.
- Hoffert, S. P. (1997). Meta-analyses gaining status in science and policymaking. *The Scientist*, 11, 1-5.
- Hogarty, K. Y. & Kromrey, J. D. (2001). We've been reporting some effect sizes: Can you guess what they mean? *Paper presented at the American Educational Research Association annual conference*, Seattle.
- Hogg, R. V. (1974). Adaptive robust procedures: A partial review and some suggestions for future applications and theory. *Journal of the American Statistical Association*, 69, 909-927.
- Huberty, C. J. (2002). A history of effect size indices. *Educational and Psychological Measurement*, 62, 227-240.
- Huberty, C. J. & Lowman, L. L. (2000). Group overlap as a basis for effect size. *Educational and Psychological Measurement*, 60, 543-563.
- Hunter, J. E. & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park: Sage.
- Huynh, C. L. (1989). A unified approach to the estimation of effect size in meta-analysis. *Paper presented at the American Educational Research Association annual conference*, San Francisco.
- Johnson, N. L. & Welch, B. L. (1939). Applications of the noncentral t-distribution. *Biometrika*, 31, 362-389.
- Kelley, T. L. (1920). Measurement of overlapping. *Journal of Educational Psychology*, 11, 458-461.

- Kelley, K. (2005). The effects of nonnormal distributions on confidence intervals around the standardized mean difference: Bootstrap and parametric confidence intervals. *Educational and Psychological Measurement*, 65, 51-69.
- Keselman, H. J., Wilcox, R. R., Lix, L. M., Algina, J. & Fradette, K. (2007). Adaptive robust estimation and testing. *British Journal of Mathematical and Statistical Psychology*, 60, 267-293.
- Klauer, K. C. (1996). Parameterschätzung. In E. Erdfelder, R. Mausfeld, T. Meiser & G. Rudinger (Hrsg.), *Handbuch Quantitative Methoden* (S. 99-107). Weinheim: Psychologie Verlags Union.
- Klein, K.-M. (1995). *Experimentelle Untersuchungen zu zwei Invarianzhypothesen des Kurzzeitgedächtnisses*. Bonn: PACE.
- Krengel, U. (2003). *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Wiesbaden: Vieweg.
- Kruskal, W. H. (1957). Historical notes on the Wilcoxon unpaired two-sample test. *Journal of the American Statistical Association*, 52, 356-360.
- Lee, A. & Gurland, J. (1977). One-sample t-test when sampling from a mixture of normal distributions. *Annals of Statistics*, 5, 803-807.
- L'Ecuyer, P. (1998). Random number generation. In J. Banks (Ed.), *The handbook of simulation* (pp. 93-137). New York: Wiley.
- Lehmann, E. L. (1993). The Fisher, Neyman-Pearson theories of testing hypotheses: One theory or two? *Journal of the American Statistical Association*, 88, 1242-1249.
- Lewis, S. & Clarke, M. (2001). Forest plots: Trying to see the wood and the trees. *British Medical Journal*, 322, 1479-1480.
- Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading: Addison-Wesley.
- Luh, W. M. & Guo, J. H. (1999). A powerful transformation trimmed mean method for one-way fixed effects ANOVA model under non-normality and inequality of variances. *British Journal of Mathematical and Statistical Psychology*, 52, 303-320.
- Lumelskii Y. P. & Volkovich, Z. E. (2007). On comparison of nonparametric and parametric approximate confidence bounds for the probability $P(X < Y)$. *Journal of Mathematical Sciences*, 146, 6016-6021.
- Mann, H. B. & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, 18, 50-60.
- Manoukian, E. B. (1986). *Modern concepts and theorems of mathematical statistics*. New York: Springer.
- Matsumoto, M. & Nishimura, T. (1998). Mersenne Twister: A 623-dimensionally equidistributed uniform pseudorandom number generator. *ACM Transactions on Modeling and Computer Simulation*, 8, 3-30.
- Mee, R. W. (1990). Confidence-intervals for probabilities and tolerance regions based on a generalization of the Mann-Whitney statistic. *Journal of the American Statistical Association*, 85, 793-800.

- McGraw, K. O. & Wong, S. P. (1992). A common language effect size statistic. *Psychological Bulletin*, *111*, 361-365.
- Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, *105*, 156-166.
- Micceri, T. (1990). Feel no guilt! Your statistics are probably robust. *Paper presented at the American Educational Research Association annual conference*, Boston.
- Moonan, W. J. (1957). Linear transformation to a set of stochastically dependent normal variables. *Journal of the American Statistical Association*, *52*, 247-252.
- Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, *11*, 364-386.
- Morris, S. B. & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, *7*, 105-125.
- Morrison, D. E. & Henkel, R. E. (1970). *The significance test controversy: A reader*. Chicago: Aldine Publishing Company.
- Munzel, U. (2006). Non-parametric confidence intervals for shift effects based on paired ranks. *Journal of Statistical Computation and Simulation*, *76*, 765-772.
- Neubert, K. (2006). *Das nichtparametrische Behrens-Fisher-Problem: Ein studentisierter Permutationstest und robuste Konfidenzintervalle für den Shift-Effekt*. Dissertation an den Mathematisch-naturwissenschaftlichen Fakultäten der Georg-August-Universität zu Göttingen.
- Neuhäuser, M., Lösch, C. & Jöckel, K. H. (2007). The Chen-Luo test in case of heteroscedasticity. *Computational Statistics and Data Analysis*, *51*, 5055-5060.
- Newcombe, R. G. (1998). Two-sided confidence intervals for the single proportion: Comparison of seven methods. *Statistics in Medicine*, *17*, 857-872.
- Newcombe, R. G. (2006a). Confidence intervals for an effect size measure based on the Mann-Whitney statistic. Part 1: General issues and tail area based methods. *Statistics in Medicine*, *25*, 543-557.
- Newcombe, R. G. (2006b). Confidence intervals for an effect size measure based on the Mann-Whitney statistic. Part 2: Asymptotic methods and evaluation. *Statistics in Medicine*, *25*, 559-573.
- Neyman, J. & Pearson, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika*, *20*, 175-240, 263-294.
- Neyman, J. & Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Transactions of the Royal Society of London, A*, *231*, 289-337.
- Oakes, M. (1986). *Statistical inference: A commentary for the social and behavioral sciences*. Chichester: Wiley.
- Owen, D. B., Craswell, K. J. & Hanson, D. L. (1964). Nonparametric upper confidence bounds for $\Pr(X < Y)$ and confidence limits for $\Pr(X < Y)$ when X and Y are normal. *Journal of the American Statistical Association*, *59*, 906-924.
- Pratt, J. W. (1968). A normal approximation for binomial, F, beta, and other common, related tail probabilities, II. *Journal of the American Statistical Association*, *63*, 1457-1483.
- Putter, J. (1955). The treatment of ties in some non-parametric tests. *Annals of Mathematical Statistics*, *26*, 368-386.

- Quenouille, M. (1949). Approximate tests of correlation in time series. *Journal of the Royal Statistical Society, B, 11*, 18-84.
- Rasch, B., Frieze, M., Hofmann, W. & Naumann, E. (2006). *Quantitative Methoden (Band 1)*. Heidelberg: Springer.
- Rinne, H. (2003). *Taschenbuch der Statistik*. Frankfurt a. M.: Verlag Harri Deutsch.
- Robey, R. R. & Barcikowski, R. S. (1992). Type I error and the number of iterations in Monte Carlo studies of robustness. *British Journal of Mathematical and Statistical Psychology, 45*, 283-288.
- Rohatgi, V. K. (1984). *Statistical inference*. New York: Wiley.
- Rosenberger, J. L. & Gasko, M. (1983). Comparing location estimators: Trimmed means, medians, and trimean. In D. C. Hoaglin, F. Mosteller, J. W. Tukey (Eds.), *Understanding robust and exploratory data analysis* (pp. 297-338). New York: Wiley.
- Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Newbury Park: Sage.
- Rosnow, R. L. & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist, 44*, 1276-1284.
- Ryu, E. & Agresti, A. (2008). Modeling and inference for an ordinal effect size measure. *Statistics in Medicine, 27*, 1703-1717.
- Sauro, J. & Lewis, J. R. (2005). Estimating completion rates from small samples using binomial confidence intervals: Comparisons and recommendations. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Orlando.
- Schenker, N. & Gentleman, J. F. (2001). On judging the significance of differences by examining the overlap between confidence intervals. *The American Statistician, 55*, 182-186.
- Schmidt, F. L. (1992). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologist, 47*, 1173-1181.
- Sen, P. K. (1967). A note on asymptotically distribution-free confidence bounds for $\Pr\{Y < X\}$ based on two independent samples. *Sankhya, A, 29*, 95-102.
- Serlin, R. C. (2000). Testing for robustness in Monte Carlo studies. *Psychological Methods, 5*, 230-240.
- Shao, J. & Tu, D. (1995). *The jackknife and bootstrap*. New York: Springer.
- Sievers, W. (1998). *Bootstrap-Inferenzverfahren*. Aachen: Shaker.
- Somers, R. H. (1962). A new asymmetric measure of association for ordinal variables. *American Sociological Review, 27*, 799-811.
- Steiger, J. H. & Fouladi, R. T. (1997). Noncentrality interval estimation and the evaluation of statistical models. In L. Harlow, S. A. Mulaik & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 221-257). Mahwah: Lawrence Erlbaum Associates.
- Stigler, S. M. (1977). Do robust estimators work with real data? *Annals of Statistics, 5*, 1055-1098.
- Stigler, S. M. (1986). *The history of statistics*. Cambridge: Belknap.
- Tatsuoka, M. (1993). Effect size. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues* (pp. 461-479). Hillsdale: Lawrence Erlbaum Associates.
- Tilton, J. W. (1937). The measurement of overlapping. *Journal of Educational Psychology, 28*, 656-662.

- Tippett, L. H. C. (1931). *The method of statistics*. London: Williams & Norgate.
- Tukey, J. W. (1958). Bias and confidence in not quite large samples. *Annals of Mathematical Statistics*, 29, 614.
- Tukey, J. W. (1977). *Exploratory data analysis*. Reading: Addison-Wesley.
- Tukey, J. W. & McLaughlin, D. H. (1963). Less vulnerable confidence and significance procedures for location based on a single sample: Trimming/winsorization I. *Sankhya, A*, 25, 331-352.
- Vargha, A. & Delaney, H. D. (2000). A critique and improvement of the CL common language effect size statistics of McGraw and Wong. *Journal of Educational and Behavioral Statistics*, 25, 101-132.
- Venables, W. (1975). Calculation of confidence intervals for non-centrality parameters. *Journal of the Royal Statistical Society, B*, 37, 406-412.
- Viechtbauer, W. (2007). Approximate confidence intervals for standardized effect sizes in the two-independent and two-dependent samples design. *Journal of Educational and Behavioral Statistics*, 32, 39-60.
- Vollset, S. E. (1993). Confidence intervals for binomial proportion. *Statistics in Medicine*, 12, 809-824.
- Westermann, R. (2000). *Wissenschaftstheorie und Experimentalmethodik*. Göttingen: Hogrefe.
- Wicherts, J. M., Borsboom, D., Kats, J. & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist*, 61, 726-728.
- Wichmann, B. A. & Hill, I. D. (2006). Generating good pseudo-random numbers. *Computational Statistics and Data Analysis*, 51, 1614-1622.
- Wilcox, R. R. (1996). *Statistics for the social sciences*. San Diego: Academic Press.
- Wilcox, R. R. (1998a). The goals and strategies of robust methods. *British Journal of Mathematical and Statistical Psychology*, 51, 1-39.
- Wilcox, R. R. (1998b). How many discoveries have been lost by ignoring modern statistical methods? *American Psychologist*, 53, 300-314.
- Wilcox, R. R. (2005). *Introduction to robust estimation and hypothesis testing*. San Diego: Elsevier Academic Press.
- Wilcox, R. R., Keselman, H. J. & Kowalchuk, R. K. (1998). Can tests for treatment group equality be improved? The bootstrap and trimmed means conjecture. *British Journal of Mathematical and Statistical Psychology*, 51, 123-134.
- Wilcox R. R. & Muska, J. (1999). Measuring effect size: A non-parametric analogue of ω^2 . *British Journal of Mathematical and Statistical Psychology*, 52, 93-110.
- Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1, 80-83.
- Wilkinson, L. & APA Task Force on Statistical Inference (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54, 594-604.
- Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22, 209-212.
- Winer, B. J., Brown, D. R. & Michels, K. M. (1991). *Statistical principles in experimental design*. London: McGraw Hill.
- Wittkowski, K. M. (1998). Versions of the sign test in the presence of ties. *Biometrics*, 54, 789-791.

- Wu, J. R. & Jiang, G. Y. (2007). Confidence intervals of effect size for paired studies. *Biometrical Journal*, 49, 765-773.
- Wu, J. R., Jiang, G. Y. & Wei, W. (2006). Confidence intervals of effect size in randomized comparative parallel-group studies. *Statistics in Medicine*, 25, 639-651.
- Yuen, K. K. (1974). The two-sample trimmed t for unequal population variances. *Biometrika*, 61, 165-170.
- Yuen, K. K. & Dixon, W. J. (1973). The approximate behaviour and performance of the two-sample trimmed t. *Biometrika*, 60, 369-374.
- Zhou, W. (2008). Statistical inference for $P(X < Y)$. *Statistics in Medicine*, 27, 257-279.
- Zhou, X. H., Li, C. M. & Yang, Z. (2008). Improving interval estimation of binomial proportions. *Philosophical Transactions of the Royal Society, A*, 366, 2405-2418.
- Zimmerman, D. W. & Zumbo, B. D. (1993). The relative power of parametric and nonparametric statistical methods. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues* (pp. 481-517). Hillsdale: Lawrence Erlbaum Associates.

Anhang A Symbolverzeichnis und Abkürzungen

$1 - \hat{\alpha}^*$	Simuliertes, aktuelles Konfidenzniveau (synonym: simulierte Überdeckungswahrscheinlichkeit)
BCA	BCA-(Bias corrected and accelerated)-Bootstrap (siehe Kapitel 2.1.2.3.2)
B_{MC}	Monte-Carlo-Schätzung des Bias (siehe Kapitel 2.1.5.1)
$ CP - .95 $	Absolute Abweichung des simulierten vom nominalen Konfidenzniveau $1 - \alpha = .95$
d	Standardisierte Lagedifferenz; ohne weitere Indizierung identisch mit der Effektgröße nach Cohen (1988); für hier berücksichtigte Varianten aus Studie 1 siehe Tabelle 4, S. 28 und aus Studie 2 siehe Tabelle 26, S. 75
$g-h$	$g-h$ -Verteilung; die Parameter und Eigenschaften der vier hier verwendeten Modelle sind in Tabelle 7, S. 32 sowie im Anhang B, Tabelle B-1 aufgeführt
Gleich	Gleichverteilung
Hyper	Negative hypergeometrische Verteilung (siehe Kapitel 2.1.4.2)
LNCP, RNCP	Left- /right-non-coverage-probability; in Studien 3 und 4 ausgewiesene Häufigkeiten, mit welchen der Parameter links von den unteren bzw. rechts von den oberen Intervallgrenzen liegt (siehe Kapitel 3.1.6)
MSE_{MC}	Monte-Carlo-Schätzung des mittleren quadratischen Fehlers (siehe Kapitel 2.1.5.1)
NC	Konfidenzintervall auf Basis der nichtzentralen t -Verteilung (siehe Kapitel 2.1.2.2 und 2.2.2.2)
N_{MC}	Anzahl der Simulationsdurchgänge je Untersuchungsbedingung; in den Studien 1 bis 4 ist $N_{MC} = 5,000$ (siehe Kapitel 2.1.4.4)
Normal	Normalverteilung
$pCP_{.936}^{.966}$, $pCP_{.936}^{1.000}$	Relative Häufigkeit, mit welcher die simulierten Konfidenzniveaus bei der gewählten nominalen Wahrscheinlichkeit von $1 - \alpha = .95$ zwischen .936 und .966 bzw. 1 liegen (zur Begründung dieser Grenzen siehe Kapitel 2.1.4.4)
Perc	Perzentil-Bootstrap (siehe Kapitel 2.1.2.3.1)
Perc-t	Perzentil- t -Bootstrap (siehe Kapitel 2.1.2.3.3)
θ	In Kapitel 2 ein interessierender Parameter; in Kapitel 3 Bezeichnung für die stochastische Überlegenheit (siehe Kapitel 3.1.2 und 3.2.1)

Anhang B Parameter der Pseudozufallszahlen

Tabelle B-1. Häufigkeitsverteilungen der simulierten, z -transformierten Verteilungen ($N = 20,000,000$)

	Normal	Gleich	$g-h-1$	$g-h-2$	$g-h-3$	$g-h-4$
$z \leq -20$	0	0	.000012	0	.000002	0
$-20 < z \leq -10$	0	0	.000137	0	.000026	0
$-10 < z \leq -9$	0	0	.000067	0	.000013	0
$-9 < z \leq -8$	0	0	.000114	0	.000026	0
$-8 < z \leq -7$	0	0	.000196	0	.000044	0
$-7 < z \leq -6$	0	0	.000357	0	.000093	0
$-6 < z \leq -5$	0	0	.000716	0	.000210	0
$-5 < z \leq -4$.000031	0	.001639	0	.000556	0
$-4 < z \leq -3$.001317	0	.004416	0	.001829	0
$-3 < z \leq -2$.021349	0	.015223	0	.008446	0
$-2 < z \leq -1$.136110	.211320	.074725	0	.067119	.019151
$-1 < z \leq 0$.341275	.288814	.402380	.630715	.486895	.587293
$0 < z \leq 1$.341278	.288640	.402560	.234773	.340150	.213865
$1 < z \leq 2$.135896	.211226	.074621	.085131	.065593	.120628
$2 < z \leq 3$.021396	0	.015167	.031082	.017064	.059063
$3 < z \leq 4$.001316	0	.004418	.011504	.006057	0
$4 < z \leq 5$.000031	0	.001648	.004262	.002581	0
$5 < z \leq 6$	0	0	.000713	.001607	.001285	0
$6 < z \leq 7$	0	0	.000361	.000588	.000707	0
$7 < z \leq 8$	0	0	.000195	.000221	.000422	0
$8 < z \leq 9$	0	0	.000111	.000076	.000262	0
$9 < z \leq 10$	0	0	.000068	.000028	.000169	0
$10 < z \leq 20$	0	0	.000143	.000014	.000396	0
$20 < z$	0	0	.000013	0	.000054	0

Tabelle B-2. Erforderliches ρ_N (Korrelation der Normalverteilungen) zur Erzeugung von stetigen Verteilungen mit ρ sowie zugehörige 20%-winsorisierte Parameter (Studien 2 und 4)

	ρ_N	ρ	cov_w	ρ_w
Normal	.4000	.400	0.147	.355
	.8000	.800	0.311	.755
Gleich	.4158	.400	0.238	.367
	.8135	.800	0.497	.766
$g-h-1$.4454	.400	0.189	.400
	.8337	.800	0.376	.794
$g-h-2$.4465	.400	0.175	.391
	.8273	.800	0.348	.777
$g-h-3$.4679	.400	0.202	.420
	.8467	.800	0.388	.809
$g-h-4$.4436	.400	0.143	.378
	.8288	.800	0.292	.770

Tabelle B-3. Erforderliche Parameter d' zur Erzeugung von stetigen Verteilungen mit gewünschter stochastischer Überlegenheit θ (Studien 3 und 4)

	$\sigma_1:\sigma_2 = 1:1$		$\sigma_1:\sigma_2 = 3:1$	
	d'	θ	d'	θ
Normal	0.000	0.500	0.000	0.500
	0.400	0.611	0.400	0.611
	0.800	0.714	0.800	0.714
	1.200	0.802	1.200	0.802
	1.600	0.871	1.600	0.871
Gleich	0.000	0.500	0.000	0.500
	0.411	0.611	0.517	0.611
	0.845	0.714	0.995	0.714
	1.284	0.802	1.403	0.802
	1.705	0.871	1.736	0.871
g-h-1	0.000	0.500	0.000	0.500
	0.295	0.611	0.275	0.611
	0.600	0.714	0.564	0.714
	0.928	0.802	0.876	0.802
	1.294	0.871	1.232	0.871
g-h-2	0.000	0.500	0.345	0.500
	0.254	0.611	0.618	0.611
	0.562	0.714	0.833	0.714
	0.927	0.802	1.001	0.802
	1.354	0.871	1.193	0.871
g-h-3	0.000	0.500	0.105	0.500
	0.272	0.611	0.355	0.611
	0.555	0.714	0.608	0.714
	0.861	0.802	0.874	0.802
	1.208	0.871	1.165	0.871
g-h-4	0.000	0.500	0.403	0.500
	0.271	0.611	0.741	0.611
	0.673	0.714	0.920	0.714
	1.136	0.802	1.118	0.802
	1.616	0.871	1.361	0.871

Tabelle B-4. Parameter der negativen hypergeometrischen Verteilungen mit gleichen Varianzen sowie zugehörige KQ- und robustifizierte Effektgrößen (Studien 1 und 2)

a_1	b_1	a_2	b_2	ρ_N	ρ	d'	d_D	d^{S_1}	d^{S_2}	d'_R	$d_{D,R}$	$d_R^{S_1}$	$d_R^{S_2}$
3.333	12.333	3.333	12.333	.000	.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3.333	12.333	3.333	12.333	.410	.400	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3.333	12.333	3.333	12.333	.815	.800	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3.333	12.333	2.517	12.670	.000	.000	0.400	0.283	0.400	0.400	0.656	0.464	0.657	0.656
3.333	12.333	2.517	12.670	.411	.400	0.400	0.365	0.400	0.400	0.656	0.576	0.657	0.656
3.333	12.333	2.517	12.670	.818	.800	0.400	0.632	0.400	0.400	0.656	0.933	0.657	0.656
3.333	12.333	1.536	12.364	.000	.000	0.800	0.566	0.800	0.800	1.338	0.946	1.340	1.337
3.333	12.333	1.536	12.364	.418	.400	0.800	0.730	0.800	0.800	1.338	1.180	1.340	1.337
3.333	12.333	1.536	12.364	.830	.800	0.800	1.265	0.800	0.800	1.338	1.949	1.340	1.337
3.333	12.333	0.696	11.462	.000	.000	1.200	0.849	1.200	1.200	2.104	1.488	2.102	2.106
3.333	12.333	0.696	11.462	.437	.400	1.200	1.095	1.200	1.200	2.104	1.879	2.102	2.106
3.333	12.333	0.696	11.462	.870	.800	1.200	1.897	1.200	1.200	2.104	3.383	2.102	2.106
3.333	12.333	0.166	10.141	.000	.000	1.600	1.131	1.600	1.600	3.650	2.581	2.933	5.434
3.333	12.333	0.166	10.141	.518	.400	1.600	1.461	1.600	1.600	3.650	3.146	2.933	5.434
3.333	12.333	0.166	10.141	.999	.800	1.600	2.339	1.600	1.600	3.650	4.657	2.933	5.434

Tabelle B-5. Parameter der negativen hypergeometrischen Verteilungen mit ungleichen Varianzen sowie zugehörige KQ- und robustifizierte Effektgrößen (Studien 1 und 2)

a_1	b_1	a_2	b_2	ρ_N	ρ	d'	d_D	d^{S_1}	d^{S_2}	d'_R	$d_{D,R}$	$d_R^{S_1}$	$d_R^{S_2}$
0.051	9.051	231.842	240.842	.000	.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.051	9.051	231.842	240.842	.489	.400	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.051	9.051	231.842	240.842	.973	.800	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.102	9.102	177.757	321.243	.000	.000	0.400	0.283	0.298	0.894	0.417	0.295	0.304	1.236
0.102	9.102	177.757	321.243	.476	.400	0.400	0.324	0.298	0.894	0.417	0.326	0.304	1.236
0.102	9.102	177.757	321.243	.946	.800	0.400	0.392	0.298	0.894	0.417	0.379	0.304	1.236
0.263	9.263	148.920	450.080	.000	.000	0.800	0.566	0.596	1.789	0.839	0.593	0.614	2.332
0.263	9.263	148.920	450.080	.455	.400	0.800	0.649	0.596	1.789	0.839	0.660	0.614	2.332
0.263	9.263	148.920	450.080	.901	.800	0.800	0.784	0.596	1.789	0.839	0.773	0.614	2.332
0.556	9.556	156.246	742.754	.000	.000	1.200	0.849	0.894	2.683	1.430	1.012	1.044	4.101
0.556	9.556	156.246	742.754	.440	.400	1.200	0.973	0.894	2.683	1.430	1.113	1.044	4.101
0.556	9.556	156.246	742.754	.873	.800	1.200	1.177	0.894	2.683	1.430	1.278	1.044	4.101
0.882	9.882	4.946	48.600	.000	.000	1.600	1.131	1.193	3.578	2.228	1.575	1.658	5.046
0.882	9.882	4.946	48.600	.433	.400	1.600	1.298	1.193	3.578	2.228	1.786	1.658	5.046
0.882	9.882	4.946	48.600	.879	.800	1.600	1.569	1.193	3.578	2.228	2.182	1.658	5.046

Tabelle B-6. Parameter der negativen hypergeometrischen Verteilungen mit gleichen Varianzen (Studien 3 und 4)

a_1	b_1	a_2	b_2	ρ_N	ρ	θ
3.333	12.333	3.333	12.333	.000	.000	.500
3.333	12.333	3.333	12.333	.410	.400	.500
3.333	12.333	3.333	12.333	.815	.800	.500
3.333	12.333	2.519	12.670	.000	.000	.611
3.333	12.333	2.519	12.670	.411	.400	.611
3.333	12.333	2.519	12.670	.818	.800	.611
3.333	12.333	1.547	12.371	.000	.000	.714
3.333	12.333	1.547	12.371	.417	.400	.714
3.333	12.333	1.547	12.371	.830	.800	.714
3.333	12.333	0.718	11.499	.000	.000	.802
3.333	12.333	0.718	11.499	.436	.400	.802
3.333	12.333	0.718	11.499	.868	.800	.802
3.333	12.333	0.205	10.293	.000	.000	.871
3.333	12.333	0.205	10.293	.502	.400	.871
3.333	12.333	0.205	10.293	.999	.800	.871

Tabelle B-7. Parameter der negativen hypergeometrischen Verteilungen mit ungleichen Varianzen (Studien 3 und 4)

a_1	b_1	a_2	b_2	ρ_N	ρ	θ
0.051	9.051	231.842	240.842	.000	.000	.500
0.051	9.051	231.842	240.842	.488	.400	.500
0.051	9.051	231.842	240.842	.973	.800	.500
0.176	9.176	8.403	38.454	.000	.000	.611
0.176	9.176	8.403	38.454	.468	.400	.611
0.176	9.176	8.403	38.454	.925	.800	.611
0.359	9.359	4.648	33.624	.000	.000	.714
0.359	9.359	4.648	33.624	.457	.400	.714
0.359	9.359	4.648	33.624	.901	.800	.714
0.556	9.556	1.923	24.496	.000	.000	.802
0.556	9.556	1.923	24.496	.466	.400	.802
0.556	9.556	1.923	24.496	.919	.800	.802
0.882	9.882	1.165	22.201	.000	.000	.871
0.882	9.882	1.165	22.201	.480	.400	.871
0.882	9.882	1.165	22.201	.960	.800	.871

Anhang C Ergebnisse der Studie 1

Tabelle C-1. \overline{B}_{MC}^2

			d_s		\hat{d}		d_s^{Boot}		\hat{d}^{Boot}	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	40	1:3	.000	.136	.000	.117	.001	.067	.002	.055
		1:1	.000	.001	.000	.000	.001	.001	.002	.002
		3:1	.	.019	.	.024	.	.031	.	.036
	120	1:3	.000	.097	.000	.092	.000	.076	.000	.072
		1:1	.000	.000	.000	.000	.000	.000	.000	.000
		3:1	.	.021	.	.023	.	.025	.	.026
Gleich	40	1:3	.000	.111	.000	.094	.001	.058	.002	.047
		1:1	.000	.000	.000	.000	.001	.001	.002	.002
		3:1	.	.022	.	.027	.	.030	.	.036
	120	1:3	.000	.090	.000	.085	.000	.075	.000	.071
		1:1	.000	.000	.000	.000	.000	.000	.000	.000
		3:1	.	.023	.	.024	.	.025	.	.027
g-h-1	40	1:3	.011	.291	.007	.259	.001	.138	.000	.118
		1:1	.012	.021	.008	.015	.001	.002	.000	.000
		3:1	.	.002	.	.005	.	.015	.	.019
	120	1:3	.003	.157	.002	.150	.000	.106	.000	.101
		1:1	.003	.005	.002	.004	.000	.001	.000	.000
		3:1	.	.011	.	.012	.	.018	.	.020
g-h-2	40	1:3	.005	.158	.002	.138	.000	.085	.001	.071
		1:1	.004	.004	.002	.003	.000	.000	.001	.001
		3:1	.	.014	.	.019	.	.026	.	.031
	120	1:3	.001	.106	.000	.101	.000	.082	.000	.078
		1:1	.001	.001	.000	.000	.000	.000	.000	.000
		3:1	.	.019	.	.021	.	.024	.	.026
g-h-3	40	1:3	.025	.306	.018	.273	.004	.153	.002	.132
		1:1	.024	.028	.017	.021	.003	.004	.001	.002
		3:1	.	.001	.	.002	.	.009	.	.013
	120	1:3	.006	.180	.005	.173	.001	.127	.001	.121
		1:1	.007	.009	.006	.007	.002	.002	.001	.002
		3:1	.	.007	.	.008	.	.014	.	.015
g-h-4	40	1:3	.001	.103	.000	.087	.001	.062	.002	.050
		1:1	.001	.001	.000	.001	.001	.001	.002	.002
		3:1	.	.023	.	.028	.	.030	.	.036
	120	1:3	.000	.087	.000	.082	.000	.075	.000	.071
		1:1	.000	.000	.000	.000	.000	.000	.000	.000
		3:1	.	.023	.	.025	.	.025	.	.027
Hyper	40	1:3	.002	.115	.001	.098	.000	.061	.001	.050
		1:1	.000	.000	.000	.000	.000	.001	.001	.002
		3:1	.	.023	.	.028	.	.030	.	.035
	120	1:3	.000	.088	.000	.084	.000	.073	.000	.069
		1:1	.000	.000	.000	.000	.000	.000	.000	.000
		3:1	.	.023	.	.025	.	.025	.	.027

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.1.4.5

Tabelle C-2. MSE_{MC}

			d_s		\hat{d}		d_s^{Boot}		\hat{d}^{Boot}	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	40	1:3	.156	.563	.149	.528	.143	.439	.138	.413
		1:1	.117	.136	.112	.130	.107	.122	.104	.118
		3:1	.	.092	.	.094	.	.098	.	.101
	120	1:3	.048	.222	.048	.215	.047	.195	.047	.189
		1:1	.039	.042	.039	.041	.038	.040	.038	.040
		3:1	.	.044	.	.046	.	.047	.	.049
Gleich	40	1:3	.147	.513	.141	.480	.134	.381	.130	.358
		1:1	.112	.120	.108	.115	.101	.105	.099	.102
		3:1	.	.088	.	.090	.	.090	.	.093
	120	1:3	.047	.207	.046	.201	.046	.184	.045	.178
		1:1	.036	.038	.036	.037	.035	.036	.035	.036
		3:1	.	.044	.	.045	.	.045	.	.047
g-h-1	40	1:3	.217	.810	.204	.758	.205	.652	.197	.612
		1:1	.180	.222	.169	.209	.171	.208	.164	.199
		3:1	.	.117	.	.115	.	.133	.	.133
	120	1:3	.074	.337	.072	.328	.076	.298	.075	.290
		1:1	.064	.077	.063	.075	.066	.080	.065	.079
		3:1	.	.053	.	.054	.	.065	.	.066
g-h-2	40	1:3	.202	.434	.192	.403	.187	.329	.181	.305
		1:1	.163	.097	.154	.092	.154	.088	.149	.085
		3:1	.	.070	.	.073	.	.079	.	.083
	120	1:3	.062	.188	.061	.182	.061	.163	.061	.157
		1:1	.050	.029	.049	.029	.050	.028	.049	.028
		3:1	.	.037	.	.038	.	.042	.	.043
g-h-3	40	1:3	.251	.743	.236	.693	.236	.575	.225	.538
		1:1	.211	.193	.197	.180	.198	.174	.188	.165
		3:1	.	.100	.	.097	.	.112	.	.112
	120	1:3	.090	.326	.087	.317	.093	.283	.091	.275
		1:1	.080	.069	.078	.067	.083	.069	.082	.068
		3:1	.	.045	.	.045	.	.056	.	.057
g-h-4	40	1:3	.158	.390	.152	.363	.142	.299	.138	.278
		1:1	.123	.084	.117	.080	.112	.073	.109	.072
		3:1	.	.070	.	.074	.	.073	.	.077
	120	1:3	.049	.169	.049	.163	.048	.152	.047	.147
		1:1	.038	.025	.037	.024	.037	.024	.037	.024
		3:1	.	.038	.	.040	.	.040	.	.041
Hyper	40	1:3	.204	.514	.195	.481	.186	.444	.179	.417
		1:1	.153	.119	.147	.114	.143	.108	.139	.105
		3:1	.	.086	.	.089	.	.088	.	.091
	120	1:3	.063	.205	.062	.199	.061	.182	.060	.177
		1:1	.049	.037	.048	.036	.048	.035	.048	.035
		3:1	.	.043	.	.045	.	.045	.	.046

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.1.4.5

Tabelle C-3. B_{MC}^2

			d'_s		\hat{d}'		d_s^{Boot}		\hat{d}^{Boot}	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	40	1:3	.001	.005	.000	.002	.001	.002	.003	.005
		1:1	.000	.001	.000	.000	.001	.001	.002	.003
		3:1	.	.000	.	.002	.	.001	.	.008
	120	1:3	.000	.000	.000	.000	.000	.000	.000	.000
		1:1	.000	.000	.000	.000	.000	.000	.000	.000
		3:1	.	.000	.	.000	.	.000	.	.001
Gleich	40	1:3	.000	.001	.000	.000	.001	.004	.003	.007
		1:1	.000	.000	.000	.000	.001	.001	.002	.004
		3:1	.	.000	.	.003	.	.000	.	.007
	120	1:3	.000	.000	.000	.000	.000	.000	.000	.001
		1:1	.000	.000	.000	.000	.000	.000	.000	.000
		3:1	.	.000	.	.000	.	.000	.	.001
g-h-1	40	1:3	.015	.057	.007	.045	.001	.004	.000	.002
		1:1	.012	.021	.007	.012	.001	.002	.000	.000
		3:1	.	.015	.	.002	.	.001	.	.001
	120	1:3	.004	.011	.002	.010	.000	.001	.000	.001
		1:1	.003	.005	.002	.003	.000	.001	.000	.000
		3:1	.	.003	.	.001	.	.000	.	.000
g-h-2	40	1:3	.004	.015	.001	.010	.001	.000	.003	.001
		1:1	.004	.004	.002	.002	.000	.000	.001	.001
		3:1	.	.003	.	.001	.	.000	.	.005
	120	1:3	.000	.002	.000	.002	.000	.000	.000	.000
		1:1	.001	.001	.000	.000	.000	.000	.000	.000
		3:1	.	.000	.	.000	.	.000	.	.001
g-h-3	40	1:3	.027	.062	.016	.049	.003	.007	.001	.004
		1:1	.024	.028	.017	.017	.003	.004	.001	.001
		3:1	.	.022	.	.005	.	.004	.	.000
	120	1:3	.007	.018	.005	.016	.001	.004	.001	.003
		1:1	.007	.009	.006	.007	.002	.002	.001	.001
		3:1	.	.006	.	.003	.	.001	.	.000
g-h-4	40	1:3	.000	.003	.000	.002	.002	.003	.004	.005
		1:1	.001	.001	.000	.001	.001	.001	.002	.004
		3:1	.	.000	.	.004	.	.000	.	.008
	120	1:3	.000	.000	.000	.000	.000	.000	.000	.001
		1:1	.000	.000	.000	.000	.000	.000	.000	.000
		3:1	.	.000	.	.000	.	.000	.	.001
Hyper	40	1:3	.001	.001	.000	.000	.001	.004	.003	.006
		1:1	.000	.000	.000	.000	.000	.001	.002	.003
		3:1	.	.000	.	.004	.	.000	.	.007
	120	1:3	.000	.000	.000	.000	.000	.000	.000	.001
		1:1	.000	.000	.000	.000	.000	.000	.000	.000
		3:1	.	.000	.	.000	.	.000	.	.001

Anmerkung: Vorpalte 1 gibt die Verteilung an, Vorpalte 2 den Stichprobenumfang N , Vorpalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.1.4.5

Tabelle C-4. MSE_{MC}

			d'_s		\hat{d}'		d_s^{Boot}		\hat{d}^{Boot}	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	40	1:3	.165	.288	.152	.274	.149	.237	.141	.231
		1:1	.117	.136	.112	.127	.107	.122	.104	.117
		3:1	.	.101	.	.090	.	.093	.	.089
	120	1:3	.050	.081	.049	.079	.048	.075	.048	.075
		1:1	.039	.042	.039	.041	.038	.040	.038	.040
		3:1	.	.032	.	.031	.	.031	.	.030
Gleich	40	1:3	.156	.259	.145	.248	.136	.194	.131	.190
		1:1	.112	.120	.108	.113	.101	.105	.099	.101
		3:1	.	.093	.	.084	.	.084	.	.080
	120	1:3	.048	.072	.047	.071	.046	.066	.046	.066
		1:1	.036	.038	.036	.037	.035	.036	.035	.036
		3:1	.	.029	.	.028	.	.029	.	.028
g-h-1	40	1:3	.227	.433	.204	.405	.212	.370	.197	.353
		1:1	.180	.222	.168	.203	.171	.208	.164	.195
		3:1	.	.168	.	.137	.	.159	.	.140
	120	1:3	.079	.138	.076	.134	.081	.136	.079	.134
		1:1	.064	.077	.063	.075	.066	.080	.066	.078
		3:1	.	.059	.	.055	.	.062	.	.060
g-h-2	40	1:3	.164	.193	.154	.181	.148	.154	.144	.149
		1:1	.163	.097	.153	.089	.154	.088	.149	.083
		3:1	.	.085	.	.071	.	.080	.	.073
	120	1:3	.050	.054	.049	.052	.049	.051	.048	.050
		1:1	.050	.029	.049	.028	.050	.028	.049	.028
		3:1	.	.026	.	.025	.	.026	.	.025
g-h-3	40	1:3	.241	.365	.217	.340	.219	.295	.205	.281
		1:1	.211	.193	.196	.173	.198	.174	.189	.161
		3:1	.	.159	.	.124	.	.147	.	.125
	120	1:3	.085	.117	.082	.114	.086	.111	.085	.109
		1:1	.080	.069	.078	.066	.083	.069	.082	.067
		3:1	.	.057	.	.052	.	.059	.	.056
g-h-4	40	1:3	.135	.180	.128	.171	.118	.140	.117	.137
		1:1	.123	.084	.117	.079	.112	.073	.109	.071
		3:1	.	.070	.	.063	.	.064	.	.062
	120	1:3	.042	.048	.041	.047	.040	.044	.040	.043
		1:1	.038	.025	.037	.024	.037	.024	.037	.023
		3:1	.	.022	.	.022	.	.022	.	.021
Hyper	40	1:3	.189	.255	.177	.244	.168	.225	.161	.220
		1:1	.153	.119	.146	.112	.143	.108	.138	.104
		3:1	.	.091	.	.082	.	.084	.	.079
	120	1:3	.058	.071	.057	.070	.056	.065	.055	.065
		1:1	.049	.037	.048	.036	.048	.035	.048	.035
		3:1	.	.029	.	.028	.	.029	.	.028

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.1.4.5

Tabelle C-5. B_{MC}^2

			$d_s^{S_1}$		\hat{d}^{S_1}		$d_s^{S_1,Boot}$		$\hat{d}^{S_1,Boot}$	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	40	1:3	.009	.008	.000	.000	.005	.005	.022	.023
		1:1	.002	.002	.000	.000	.001	.001	.004	.004
		3:1	.	.001	.	.000	.	.000	.	.002
	120	1:3	.001	.001	.000	.000	.000	.000	.002	.002
		1:1	.000	.000	.000	.000	.000	.000	.000	.001
		3:1	.	.000	.	.000	.	.000	.	.000
Gleich	40	1:3	.003	.002	.001	.002	.007	.008	.025	.027
		1:1	.000	.000	.000	.001	.001	.001	.004	.005
		3:1	.	.000	.	.000	.	.000	.	.002
	120	1:3	.000	.000	.000	.000	.000	.000	.002	.002
		1:1	.000	.000	.000	.000	.000	.000	.001	.000
		3:1	.	.000	.	.000	.	.000	.	.000
g-h-1	40	1:3	.094	.098	.038	.041	.002	.002	.002	.002
		1:1	.033	.031	.018	.017	.003	.002	.000	.000
		3:1	.	.019	.	.012	.	.003	.	.001
	120	1:3	.019	.018	.012	.011	.002	.002	.000	.000
		1:1	.007	.007	.005	.005	.001	.001	.000	.000
		3:1	.	.004	.	.003	.	.001	.	.000
g-h-2	40	1:3	.036	.029	.010	.009	.000	.000	.009	.009
		1:1	.009	.007	.003	.003	.000	.000	.001	.001
		3:1	.	.003	.	.001	.	.000	.	.001
	120	1:3	.004	.004	.001	.001	.000	.000	.001	.001
		1:1	.001	.001	.000	.000	.000	.000	.000	.000
		3:1	.	.000	.	.000	.	.000	.	.000
g-h-3	40	1:3	.121	.104	.055	.045	.008	.005	.000	.001
		1:1	.045	.040	.027	.023	.008	.006	.002	.001
		3:1	.	.028	.	.019	.	.006	.	.003
	120	1:3	.027	.026	.018	.018	.006	.006	.002	.003
		1:1	.013	.012	.010	.009	.003	.003	.002	.002
		3:1	.	.007	.	.006	.	.002	.	.001
g-h-4	40	1:3	.005	.006	.003	.006	.007	.005	.025	.021
		1:1	.001	.001	.001	.001	.001	.001	.004	.005
		3:1	.	.000	.	.001	.	.000	.	.002
	120	1:3	.000	.000	.000	.001	.000	.000	.002	.002
		1:1	.000	.000	.000	.000	.000	.000	.001	.000
		3:1	.	.000	.	.000	.	.000	.	.000
Hyper	40	1:3	.004	.002	.001	.002	.003	.006	.019	.024
		1:1	.000	.000	.000	.001	.001	.001	.004	.005
		3:1	.	.000	.	.000	.	.000	.	.002
	120	1:3	.000	.000	.000	.000	.000	.000	.002	.002
		1:1	.000	.000	.000	.000	.000	.000	.000	.000
		3:1	.	.000	.	.000	.	.000	.	.000

Anmerkung: Vorpalte 1 gibt die Verteilung an, Vorpalte 2 den Stichprobenumfang N , Vorpalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.1.4.5

Tabelle C-6. MSE_{MC}

			$d_s^{S_1}$		\hat{d}^{S_1}		$d_s^{S_1,Boot}$		$\hat{d}^{S_1,Boot}$	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	40	1:3	.268	.333	.216	.272	.200	.249	.185	.226
		1:1	.142	.147	.129	.134	.124	.128	.118	.121
		3:1	.	.106	.	.100	.	.097	.	.094
	120	1:3	.066	.086	.062	.081	.060	.079	.059	.076
		1:1	.045	.044	.043	.042	.043	.042	.042	.041
		3:1	.	.033	.	.032	.	.032	.	.032
Gleich	40	1:3	.221	.285	.184	.239	.152	.189	.147	.179
		1:1	.123	.125	.113	.115	.107	.106	.102	.101
		3:1	.	.095	.	.090	.	.086	.	.083
	120	1:3	.055	.075	.053	.071	.050	.068	.049	.066
		1:1	.038	.038	.037	.038	.037	.037	.036	.036
		3:1	.	.030	.	.029	.	.029	.	.029
g-h-1	40	1:3	.543	.613	.414	.471	.396	.454	.332	.380
		1:1	.287	.268	.253	.235	.249	.240	.227	.219
		3:1	.	.193	.	.177	.	.180	.	.169
	120	1:3	.154	.168	.140	.153	.145	.160	.136	.150
		1:1	.091	.089	.087	.085	.091	.091	.088	.088
		3:1	.	.066	.	.064	.	.069	.	.067
g-h-2	40	1:3	.269	.255	.206	.199	.200	.181	.177	.160
		1:1	.150	.108	.134	.095	.130	.093	.121	.087
		3:1	.	.083	.	.077	.	.076	.	.072
	120	1:3	.061	.061	.056	.056	.058	.056	.056	.054
		1:1	.041	.030	.039	.029	.040	.030	.039	.029
		3:1	.	.025	.	.024	.	.025	.	.024
g-h-3	40	1:3	.537	.509	.403	.383	.372	.359	.305	.296
		1:1	.277	.230	.241	.199	.232	.197	.209	.177
		3:1	.	.174	.	.158	.	.157	.	.146
	120	1:3	.145	.145	.130	.130	.132	.132	.122	.122
		1:1	.091	.079	.086	.074	.089	.077	.086	.074
		3:1	.	.061	.	.059	.	.063	.	.061
g-h-4	40	1:3	.168	.217	.138	.183	.140	.150	.137	.142
		1:1	.101	.085	.093	.079	.086	.071	.083	.069
		3:1	.	.067	.	.063	.	.060	.	.058
	120	1:3	.038	.048	.036	.046	.034	.043	.034	.042
		1:1	.029	.024	.028	.024	.028	.023	.027	.023
		3:1	.	.021	.	.021	.	.020	.	.020
Hyper	40	1:3	.236	.284	.195	.238	.182	.214	.168	.198
		1:1	.133	.121	.123	.112	.118	.107	.112	.102
		3:1	.	.090	.	.085	.	.082	.	.080
	120	1:3	.062	.072	.058	.069	.056	.065	.055	.064
		1:1	.040	.037	.039	.036	.039	.035	.038	.035
		3:1	.	.029	.	.028	.	.028	.	.028

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.1.4.5

Tabelle C-7. B_{MC}^2

			$d_s^{S_2}$		\hat{d}^{S_2}		$d_s^{S_2,Boot}$		$\hat{d}^{S_2,Boot}$	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	40	1:3	.001	.001	.000	.000	.000	.000	.002	.002
		1:1	.002	.002	.000	.000	.001	.001	.004	.004
		3:1	.	.008	.	.000	.	.006	.	.024
	120	1:3	.000	.000	.000	.000	.000	.000	.000	.000
		1:1	.000	.000	.000	.000	.000	.000	.000	.000
		3:1	.	.001	.	.000	.	.000	.	.002
Gleich	40	1:3	.000	.000	.000	.000	.000	.001	.002	.002
		1:1	.000	.000	.000	.000	.001	.001	.005	.004
		3:1	.	.002	.	.002	.	.007	.	.026
	120	1:3	.000	.000	.000	.000	.000	.000	.000	.000
		1:1	.000	.000	.000	.000	.000	.000	.000	.000
		3:1	.	.000	.	.000	.	.000	.	.002
g-h-1	40	1:3	.017	.019	.010	.012	.002	.002	.000	.001
		1:1	.033	.034	.018	.019	.003	.003	.000	.000
		3:1	.	.095	.	.039	.	.002	.	.002
	120	1:3	.004	.004	.003	.003	.001	.000	.000	.000
		1:1	.007	.007	.005	.005	.001	.001	.000	.000
		3:1	.	.018	.	.011	.	.002	.	.000
g-h-2	40	1:3	.014	.013	.008	.007	.000	.000	.000	.000
		1:1	.031	.024	.017	.012	.000	.000	.002	.001
		3:1	.	.114	.	.051	.	.000	.	.009
	120	1:3	.002	.001	.001	.001	.000	.000	.000	.000
		1:1	.004	.003	.002	.002	.000	.000	.000	.000
		3:1	.	.011	.	.006	.	.000	.	.000
g-h-3	40	1:3	.045	.039	.033	.028	.009	.008	.005	.004
		1:1	.074	.065	.049	.043	.012	.011	.005	.004
		3:1	.	.175	.	.089	.	.011	.	.000
	120	1:3	.011	.009	.009	.008	.003	.002	.002	.001
		1:1	.019	.016	.015	.013	.005	.004	.003	.002
		3:1	.	.039	.	.028	.	.008	.	.004
g-h-4	40	1:3	.002	.001	.000	.000	.000	.000	.002	.002
		1:1	.006	.003	.001	.000	.002	.002	.006	.007
		3:1	.	.025	.	.004	.	.019	.	.044
	120	1:3	.000	.000	.000	.000	.000	.000	.000	.000
		1:1	.000	.000	.000	.000	.000	.000	.001	.000
		3:1	.	.001	.	.000	.	.000	.	.002
Hyper	40	1:3	.006	.001	.002	.000	.000	.000	.001	.002
		1:1	.012	.003	.004	.000	.001	.000	.003	.003
		3:1	.	.011	.	.000	.	.001	.	.010
	120	1:3	.001	.000	.000	.000	.000	.000	.000	.000
		1:1	.001	.000	.000	.000	.000	.000	.000	.000
		3:1	.	.001	.	.000	.	.000	.	.001

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.1.4.5

Tabelle C-8. MSE_{MC}

			$d_s^{S_2}$		\hat{d}^{S_2}		$d_s^{S_2,Boot}$		$\hat{d}^{S_2,Boot}$	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	40	1:3	.162	.223	.153	.211	.149	.204	.142	.195
		1:1	.141	.146	.129	.133	.123	.127	.117	.121
		3:1	.	.197	.	.158	.	.148	.	.142
	120	1:3	.050	.067	.049	.066	.049	.065	.048	.065
		1:1	.045	.043	.044	.042	.043	.041	.042	.041
		3:1	.	.048	.	.044	.	.043	.	.043
Gleich	40	1:3	.147	.196	.140	.186	.137	.185	.131	.177
		1:1	.123	.121	.113	.112	.107	.108	.102	.103
		3:1	.	.139	.	.116	.	.105	.	.108
	120	1:3	.047	.065	.047	.064	.046	.064	.046	.063
		1:1	.038	.038	.037	.037	.037	.036	.036	.036
		3:1	.	.036	.	.034	.	.033	.	.033
g-h-1	40	1:3	.267	.361	.248	.336	.244	.320	.230	.301
		1:1	.286	.305	.251	.269	.249	.257	.227	.234
		3:1	.	.516	.	.391	.	.364	.	.305
	120	1:3	.086	.109	.084	.106	.086	.107	.084	.105
		1:1	.091	.092	.087	.088	.091	.091	.088	.088
		3:1	.	.136	.	.123	.	.126	.	.118
g-h-2	40	1:3	.327	.352	.305	.329	.287	.307	.273	.291
		1:1	.412	.331	.368	.296	.347	.280	.322	.260
		3:1	.	.787	.	.614	.	.629	.	.535
	120	1:3	.091	.099	.089	.096	.087	.094	.086	.093
		1:1	.105	.089	.102	.085	.099	.084	.097	.081
		3:1	.	.148	.	.136	.	.134	.	.127
g-h-3	40	1:3	.387	.477	.357	.444	.351	.425	.328	.399
		1:1	.442	.419	.388	.369	.380	.354	.344	.320
		3:1	.	.729	.	.552	.	.507	.	.414
	120	1:3	.127	.142	.123	.138	.129	.141	.126	.138
		1:1	.145	.133	.138	.127	.145	.131	.139	.127
		3:1	.	.204	.	.184	.	.187	.	.174
g-h-4	40	1:3	.206	.244	.194	.230	.178	.216	.170	.206
		1:1	.236	.183	.213	.166	.183	.149	.173	.142
		3:1	.	.458	.	.366	.	.385	.	.350
	120	1:3	.061	.075	.060	.073	.058	.072	.057	.071
		1:1	.060	.050	.058	.048	.055	.047	.054	.046
		3:1	.	.069	.	.064	.	.059	.	.057
Hyper	40	1:3	.265	.223	.248	.210	.237	.240	.226	.229
		1:1	.342	.166	.309	.151	.309	.154	.288	.145
		3:1	.	.228	.	.182	.	.207	.	.182
	120	1:3	.073	.073	.072	.072	.069	.072	.068	.071
		1:1	.079	.050	.076	.049	.072	.048	.071	.047
		3:1	.	.063	.	.059	.	.059	.	.057

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.1.4.5

Tabelle C-9. B_{MC}^2

			$d_{s,R}$		\hat{d}_R		$d_{s,R}^{Boot}$		\hat{d}_R^{Boot}	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	40	1:3	.004	.225	.001	.180	.002	.161	.000	.125
		1:1	.004	.009	.001	.003	.002	.002	.000	.000
		3:1	.	.011	.	.018	.	.018	.	.026
	120	1:3	.000	.118	.000	.109	.000	.106	.000	.097
		1:1	.000	.001	.000	.000	.000	.000	.000	.000
		3:1	.	.018	.	.021	.	.021	.	.023
Gleich	40	1:3	.006	.252	.002	.203	.001	.138	.000	.106
		1:1	.006	.009	.002	.004	.001	.000	.000	.000
		3:1	.	.010	.	.016	.	.021	.	.030
	120	1:3	.000	.124	.000	.114	.000	.099	.000	.090
		1:1	.001	.001	.000	.000	.000	.000	.000	.000
		3:1	.	.018	.	.020	.	.022	.	.024
g-h-1	40	1:3	.002	.204	.000	.162	.003	.183	.000	.144
		1:1	.002	.006	.000	.001	.003	.003	.000	.000
		3:1	.	.011	.	.019	.	.015	.	.023
	120	1:3	.000	.110	.000	.101	.000	.110	.000	.101
		1:1	.000	.000	.000	.000	.000	.000	.000	.000
		3:1	.	.019	.	.021	.	.020	.	.022
g-h-2	40	1:3	.009	.177	.003	.143	.002	.119	.000	.093
		1:1	.004	.006	.001	.002	.002	.002	.000	.000
		3:1	.	.009	.	.014	.	.012	.	.017
	120	1:3	.001	.086	.000	.079	.000	.076	.000	.070
		1:1	.000	.001	.000	.000	.000	.000	.000	.000
		3:1	.	.013	.	.015	.	.014	.	.016
g-h-3	40	1:3	.003	.180	.000	.143	.003	.163	.001	.128
		1:1	.002	.006	.000	.002	.003	.003	.000	.001
		3:1	.	.010	.	.017	.	.012	.	.019
	120	1:3	.000	.099	.000	.090	.000	.099	.000	.090
		1:1	.000	.001	.000	.000	.001	.000	.000	.000
		3:1	.	.017	.	.019	.	.017	.	.019
g-h-4	40	1:3	.015	.181	.007	.147	.001	.095	.000	.072
		1:1	.007	.007	.002	.003	.001	.001	.000	.000
		3:1	.	.009	.	.014	.	.013	.	.019
	120	1:3	.001	.089	.001	.082	.000	.072	.000	.066
		1:1	.000	.001	.000	.000	.000	.000	.000	.000
		3:1	.	.013	.	.015	.	.015	.	.016
Hyper	40	1:3	.038	.301	.023	.245	.048	.137	.032	.104
		1:1	.001	.004	.003	.002	.000	.003	.001	.007
		3:1	.	.028	.	.038	.	.043	.	.056
	120	1:3	.017	.118	.014	.108	.016	.081	.013	.074
		1:1	.001	.002	.002	.002	.001	.003	.002	.004
		3:1	.	.039	.	.042	.	.043	.	.047

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.1.4.5

Tabelle C-10. MSE_{MC}

			$d_{s,R}$		\hat{d}_R		$d_{s,R}^{Boot}$		\hat{d}_R^{Boot}	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	40	1:3	.215	.866	.198	.778	.231	.885	.213	.800
		1:1	.170	.214	.156	.194	.183	.223	.169	.206
		3:1	.	.117	.	.117	.	.130	.	.131
	120	1:3	.062	.287	.060	.274	.064	.286	.063	.273
		1:1	.052	.057	.050	.055	.054	.059	.052	.058
		3:1	.	.049	.	.051	.	.052	.	.054
Gleich	40	1:3	.307	1.195	.282	1.082	.308	1.102	.286	1.004
		1:1	.248	.289	.227	.264	.239	.264	.222	.246
		3:1	.	.154	.	.151	.	.160	.	.160
	120	1:3	.090	.364	.088	.349	.091	.336	.089	.323
		1:1	.071	.077	.069	.075	.071	.076	.069	.074
		3:1	.	.060	.	.062	.	.064	.	.065
g-h-1	40	1:3	.128	.601	.118	.532	.154	.709	.141	.634
		1:1	.105	.143	.096	.129	.129	.173	.118	.159
		3:1	.	.083	.	.086	.	.099	.	.102
	120	1:3	.036	.219	.035	.207	.040	.238	.039	.226
		1:1	.031	.039	.030	.038	.035	.044	.033	.043
		3:1	.	.040	.	.042	.	.043	.	.045
g-h-2	40	1:3	.197	.513	.179	.457	.215	.515	.199	.462
		1:1	.154	.111	.140	.100	.176	.122	.162	.112
		3:1	.	.068	.	.069	.	.077	.	.079
	120	1:3	.054	.170	.053	.161	.058	.171	.057	.163
		1:1	.044	.030	.043	.029	.049	.032	.047	.032
		3:1	.	.030	.	.032	.	.032	.	.034
g-h-3	40	1:3	.122	.489	.112	.431	.149	.576	.137	.514
		1:1	.101	.110	.093	.099	.128	.136	.117	.124
		3:1	.	.067	.	.070	.	.081	.	.083
	120	1:3	.035	.179	.034	.169	.039	.195	.038	.184
		1:1	.030	.030	.029	.029	.034	.035	.033	.034
		3:1	.	.033	.	.035	.	.035	.	.037
g-h-4	40	1:3	.292	.698	.265	.628	.287	.621	.267	.563
		1:1	.219	.161	.200	.147	.231	.157	.214	.146
		3:1	.	.092	.	.091	.	.096	.	.096
	120	1:3	.078	.217	.076	.207	.080	.204	.078	.195
		1:1	.061	.041	.059	.040	.064	.042	.062	.041
		3:1	.	.038	.	.039	.	.040	.	.041
Hyper	40	1:3	.464	2.195	.420	2.011	.658	2.654	.602	2.451
		1:1	.253	.425	.238	.394	.322	.405	.301	.382
		3:1	.	.189	.	.189	.	.179	.	.182
	120	1:3	.153	.462	.148	.445	.192	.384	.186	.370
		1:1	.079	.094	.078	.092	.096	.092	.094	.091
		3:1	.	.084	.	.087	.	.090	.	.093

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.1.4.5

Tabelle C-11. B_{MC}^2

			$d'_{s,R}$		\hat{d}'_R		$d'^{Boot}_{s,R}$		\hat{d}'^{Boot}_R	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	40	1:3	.007	.036	.000	.021	.004	.004	.000	.001
		1:1	.004	.009	.000	.001	.002	.002	.000	.000
		3:1	.	.004	.	.005	.	.001	.	.009
	120	1:3	.001	.003	.000	.002	.000	.001	.000	.000
		1:1	.000	.001	.000	.000	.000	.000	.000	.000
		3:1	.	.000	.	.000	.	.000	.	.001
Gleich	40	1:3	.011	.045	.001	.028	.002	.001	.000	.000
		1:1	.006	.009	.002	.001	.001	.000	.000	.002
		3:1	.	.005	.	.004	.	.000	.	.012
	120	1:3	.001	.004	.000	.002	.000	.000	.000	.000
		1:1	.001	.001	.000	.000	.000	.000	.000	.000
		3:1	.	.001	.	.000	.	.000	.	.001
g-h-1	40	1:3	.004	.032	.000	.018	.005	.010	.000	.005
		1:1	.002	.006	.000	.000	.003	.003	.000	.000
		3:1	.	.003	.	.006	.	.003	.	.007
	120	1:3	.000	.002	.000	.001	.001	.001	.000	.000
		1:1	.000	.000	.000	.000	.000	.000	.000	.000
		3:1	.	.000	.	.000	.	.000	.	.000
g-h-2	40	1:3	.009	.033	.001	.020	.003	.006	.000	.002
		1:1	.004	.006	.001	.001	.002	.002	.000	.000
		3:1	.	.002	.	.005	.	.001	.	.006
	120	1:3	.001	.003	.000	.002	.000	.001	.000	.000
		1:1	.000	.001	.000	.000	.000	.000	.000	.000
		3:1	.	.000	.	.000	.	.000	.	.000
g-h-3	40	1:3	.004	.027	.000	.015	.005	.009	.000	.004
		1:1	.002	.006	.000	.000	.003	.003	.000	.000
		3:1	.	.002	.	.005	.	.003	.	.006
	120	1:3	.000	.002	.000	.001	.001	.001	.000	.000
		1:1	.000	.001	.000	.000	.001	.000	.000	.000
		3:1	.	.000	.	.000	.	.000	.	.000
g-h-4	40	1:3	.015	.033	.003	.020	.002	.002	.000	.001
		1:1	.007	.007	.002	.002	.001	.001	.000	.000
		3:1	.	.003	.	.005	.	.001	.	.007
	120	1:3	.001	.003	.000	.002	.000	.000	.000	.000
		1:1	.000	.001	.000	.000	.000	.000	.000	.000
		3:1	.	.000	.	.000	.	.000	.	.000
Hyper	40	1:3	.006	.041	.000	.025	.004	.000	.001	.002
		1:1	.001	.004	.005	.002	.000	.003	.002	.012
		3:1	.	.001	.	.018	.	.002	.	.030
	120	1:3	.000	.002	.001	.002	.000	.003	.001	.004
		1:1	.001	.002	.002	.003	.001	.003	.002	.005
		3:1	.	.002	.	.006	.	.003	.	.008

Anmerkung: Vorpalte 1 gibt die Verteilung an, Vorpalte 2 den Stichprobenumfang N , Vorpalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.1.4.5

Tabelle C-12. MSE_{MC}

			$d'_{s,R}$		\hat{d}'_R		$d'^{Boot}_{s,R}$		\hat{d}'^{Boot}_R	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	40	1:3	.240	.531	.199	.475	.266	.575	.224	.535
		1:1	.170	.214	.154	.187	.183	.223	.169	.201
		3:1	.	.149	.	.118	.	.157	.	.131
	120	1:3	.066	.118	.063	.115	.070	.124	.067	.121
		1:1	.052	.057	.050	.055	.054	.059	.052	.057
		3:1	.	.042	.	.039	.	.044	.	.041
Gleich	40	1:3	.348	.773	.288	.697	.348	.738	.298	.694
		1:1	.248	.289	.225	.254	.239	.264	.222	.240
		3:1	.	.205	.	.159	.	.195	.	.162
	120	1:3	.096	.162	.092	.157	.096	.153	.093	.150
		1:1	.071	.077	.069	.074	.071	.076	.069	.073
		3:1	.	.059	.	.054	.	.058	.	.055
g-h-1	40	1:3	.144	.369	.119	.326	.187	.475	.154	.437
		1:1	.105	.143	.095	.126	.129	.173	.118	.156
		3:1	.	.098	.	.081	.	.117	.	.098
	120	1:3	.041	.083	.039	.081	.047	.098	.045	.095
		1:1	.031	.039	.030	.038	.035	.044	.033	.042
		3:1	.	.029	.	.027	.	.032	.	.030
g-h-2	40	1:3	.177	.289	.151	.254	.201	.320	.179	.295
		1:1	.154	.111	.138	.096	.176	.122	.162	.109
		3:1	.	.086	.	.068	.	.096	.	.078
	120	1:3	.046	.060	.045	.058	.051	.066	.050	.064
		1:1	.044	.030	.043	.028	.049	.032	.047	.031
		3:1	.	.024	.	.022	.	.026	.	.024
g-h-3	40	1:3	.131	.284	.109	.251	.171	.369	.142	.340
		1:1	.101	.110	.092	.096	.128	.136	.117	.122
		3:1	.	.079	.	.066	.	.098	.	.080
	120	1:3	.036	.061	.034	.059	.042	.073	.040	.071
		1:1	.030	.030	.029	.029	.034	.035	.033	.033
		3:1	.	.022	.	.021	.	.025	.	.024
g-h-4	40	1:3	.254	.405	.219	.359	.254	.386	.231	.358
		1:1	.219	.161	.196	.140	.231	.157	.213	.141
		3:1	.	.120	.	.093	.	.121	.	.098
	120	1:3	.066	.084	.063	.081	.068	.084	.066	.081
		1:1	.061	.041	.059	.039	.064	.042	.062	.040
		3:1	.	.035	.	.033	.	.037	.	.034
Hyper	40	1:3	.381	1.576	.334	1.400	.527	2.159	.469	1.929
		1:1	.253	.425	.235	.378	.322	.405	.297	.373
		3:1	.	.231	.	.190	.	.201	.	.176
	120	1:3	.105	.206	.102	.201	.129	.174	.126	.171
		1:1	.079	.094	.078	.091	.096	.092	.094	.091
		3:1	.	.068	.	.067	.	.071	.	.070

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.1.4.5

Tabelle C-13. B_{MC}^2

			$d_{s,R}^{S_1}$		$\hat{d}_R^{S_1}$		$d_{s,R}^{S_1,Boot}$		$\hat{d}_R^{S_1,Boot}$	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	40	1:3	.080	.072	.007	.005	.268	.405	.348	.476
		1:1	.012	.013	.001	.001	.001	.001	.002	.002
		3:1	.	.005	.	.001	.	.001	.	.000
	120	1:3	.005	.005	.000	.001	.001	.001	.000	.001
		1:1	.001	.001	.000	.000	.000	.000	.000	.000
		3:1	.	.001	.	.000	.	.000	.	.000
Gleich	40	1:3	.094	.083	.011	.008	.407	.519	.476	.578
		1:1	.014	.013	.002	.001	.000	.000	.004	.005
		3:1	.	.005	.	.001	.	.000	.	.001
	120	1:3	.005	.005	.001	.001	.000	.000	.001	.001
		1:1	.001	.001	.000	.000	.000	.000	.000	.000
		3:1	.	.001	.	.000	.	.000	.	.000
g-h-1	40	1:3	.069	.073	.004	.005	.282	.283	.362	.436
		1:1	.011	.010	.001	.001	.001	.001	.001	.001
		3:1	.	.004	.	.000	.	.001	.	.000
	120	1:3	.005	.004	.000	.000	.001	.001	.000	.000
		1:1	.001	.001	.000	.000	.000	.000	.000	.000
		3:1	.	.000	.	.000	.	.000	.	.000
g-h-2	40	1:3	.096	.070	.013	.010	.430	.258	.487	.306
		1:1	.014	.010	.002	.002	.001	.001	.001	.001
		3:1	.	.003	.	.001	.	.001	.	.000
	120	1:3	.006	.005	.001	.001	.001	.001	.000	.000
		1:1	.001	.001	.000	.000	.000	.000	.000	.000
		3:1	.	.000	.	.000	.	.000	.	.000
g-h-3	40	1:3	.069	.059	.004	.004	.294	.212	.445	.283
		1:1	.009	.010	.000	.001	.001	.002	.001	.001
		3:1	.	.004	.	.000	.	.001	.	.000
	120	1:3	.004	.004	.000	.000	.001	.001	.000	.000
		1:1	.001	.001	.000	.000	.001	.000	.000	.000
		3:1	.	.000	.	.000	.	.000	.	.000
g-h-4	40	1:3	.110	.070	.018	.014	.651	.516	.685	.578
		1:1	.014	.011	.003	.003	.001	.001	.003	.002
		3:1	.	.004	.	.001	.	.000	.	.000
	120	1:3	.007	.005	.001	.001	.000	.000	.001	.000
		1:1	.001	.001	.000	.000	.000	.000	.000	.000
		3:1	.	.000	.	.000	.	.000	.	.000
Hyper	40	1:3	.076	.070	.002	.006	.016	.012	.006	.063
		1:1	.003	.005	.001	.002	.000	.007	.008	.021
		3:1	.	.002	.	.003	.	.004	.	.010
	120	1:3	.002	.002	.000	.003	.000	.004	.004	.011
		1:1	.000	.002	.001	.004	.001	.004	.003	.007
		3:1	.	.003	.	.004	.	.004	.	.006

Anmerkung: Vorpalte 1 gibt die Verteilung an, Vorpalte 2 den Stichprobenumfang N , Vorpalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.1.4.5

Tabelle C-14. MSE_{MC}

			$d_{s,R}^{S_1}$		$\hat{d}_R^{S_1}$		$d_{s,R}^{S_1,Boot}$		$\hat{d}_R^{S_1,Boot}$	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	40	1:3	.767	.913	.494	.602	16.867	77.022	12.119	54.808
		1:1	.249	.253	.206	.209	.247	.243	.215	.212
		3:1	.	.162	.	.144	.	.164	.	.149
	120	1:3	.111	.136	.097	.120	.114	.137	.104	.125
		1:1	.064	.062	.060	.058	.065	.063	.062	.061
		3:1	.	.044	.	.043	.	.046	.	.044
Gleich	40	1:3	1.066	1.174	.700	.781	53.750	42.803	38.304	30.564
		1:1	.330	.324	.275	.271	.298	.275	.261	.242
		3:1	.	.214	.	.191	.	.199	.	.183
	120	1:3	.139	.178	.123	.159	.131	.162	.120	.149
		1:1	.081	.082	.076	.077	.079	.079	.076	.075
		3:1	.	.060	.	.058	.	.060	.	.058
g-h-1	40	1:3	.608	.688	.387	.441	44.585	32.679	31.779	55.259
		1:1	.185	.183	.152	.150	.199	.202	.173	.175
		3:1	.	.113	.	.099	.	.122	.	.111
	120	1:3	.093	.103	.081	.091	.105	.116	.095	.105
		1:1	.045	.045	.042	.043	.049	.050	.047	.047
		3:1	.	.032	.	.030	.	.034	.	.033
g-h-2	40	1:3	.571	.520	.350	.330	88.603	43.538	63.015	30.998
		1:1	.181	.134	.147	.109	.193	.136	.167	.118
		3:1	.	.089	.	.079	.	.093	.	.085
	120	1:3	.079	.072	.067	.062	.089	.076	.081	.069
		1:1	.043	.032	.040	.030	.048	.035	.045	.033
		3:1	.	.024	.	.023	.	.026	.	.025
g-h-3	40	1:3	.554	.511	.348	.324	23.891	25.253	58.113	18.040
		1:1	.159	.143	.130	.116	.177	.159	.154	.137
		3:1	.	.089	.	.078	.	.100	.	.090
	120	1:3	.078	.075	.068	.066	.090	.086	.081	.078
		1:1	.040	.035	.038	.033	.045	.039	.043	.037
		3:1	.	.025	.	.024	.	.027	.	.026
g-h-4	40	1:3	.791	.780	.501	.517	90.458	108.079	64.371	117.291
		1:1	.226	.195	.186	.162	.217	.167	.190	.145
		3:1	.	.123	.	.110	.	.116	.	.106
	120	1:3	.087	.095	.075	.083	.093	.089	.085	.081
		1:1	.052	.042	.049	.039	.055	.042	.052	.040
		3:1	.	.034	.	.033	.	.035	.	.034
Hyper	40	1:3	.843	3.407	.546	2.372	1.267	5.582	.893	4.013
		1:1	.289	.614	.248	.530	.348	.768	.309	.680
		3:1	.	.259	.	.238	.	.237	.	.223
	120	1:3	.141	.228	.127	.209	.176	.191	.164	.181
		1:1	.075	.095	.073	.093	.096	.094	.094	.092
		3:1	.	.069	.	.068	.	.072	.	.071

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.1.4.5

Tabelle C-15. \overline{B}_{MC}^2

			$d_{s,R}^{S_2}$		$\hat{d}_R^{S_2}$		$d_{s,R}^{S_2,Boot}$		$\hat{d}_R^{S_2,Boot}$	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	40	1:3	.005	.005	.001	.001	.001	.001	.000	.000
		1:1	.012	.013	.001	.001	.001	.001	.002	.002
		3:1	.	.078	.	.006	.	.593	.	.732
	120	1:3	.000	.000	.000	.000	.000	.000	.000	.000
		1:1	.001	.001	.000	.000	.000	.000	.000	.000
		3:1	.	.005	.	.001	.	.001	.	.000
Gleich	40	1:3	.006	.004	.001	.000	.000	.000	.001	.002
		1:1	.014	.014	.002	.002	.000	.000	.004	.004
		3:1	.	.087	.	.009	.	.442	.	.508
	120	1:3	.000	.000	.000	.000	.000	.000	.000	.000
		1:1	.001	.001	.000	.000	.000	.000	.000	.000
		3:1	.	.006	.	.001	.	.000	.	.001
g-h-1	40	1:3	.004	.004	.000	.000	.001	.001	.000	.000
		1:1	.012	.011	.001	.001	.001	.001	.001	.001
		3:1	.	.073	.	.005	.	.247	.	.328
	120	1:3	.000	.000	.000	.000	.000	.000	.000	.000
		1:1	.001	.001	.000	.000	.000	.000	.000	.000
		3:1	.	.004	.	.000	.	.001	.	.000
g-h-2	40	1:3	.021	.021	.009	.011	.000	.000	.001	.001
		1:1	.035	.024	.012	.008	.000	.000	.004	.003
		3:1	.	.133	.	.032	.	.476	.	.503
	120	1:3	.002	.002	.001	.001	.000	.000	.000	.000
		1:1	.003	.002	.001	.001	.000	.000	.000	.000
		3:1	.	.008	.	.002	.	.000	.	.000
g-h-3	40	1:3	.007	.006	.001	.001	.001	.001	.000	.000
		1:1	.015	.014	.002	.002	.002	.002	.001	.001
		3:1	.	.076	.	.008	.	.279	.	.347
	120	1:3	.001	.001	.000	.000	.000	.000	.000	.000
		1:1	.002	.001	.000	.000	.001	.000	.000	.000
		3:1	.	.004	.	.000	.	.001	.	.000
g-h-4	40	1:3	.030	.026	.015	.015	.000	.000	.001	.001
		1:1	.050	.032	.020	.012	.001	.001	.008	.007
		3:1	.	.180	.	.053	.	1.440	.	1.310
	120	1:3	.003	.003	.001	.002	.000	.000	.000	.000
		1:1	.003	.003	.001	.001	.000	.000	.000	.000
		3:1	.	.010	.	.004	.	.000	.	.001
Hyper	40	1:3	.014	.024	.004	.011	.018	.012	.009	.005
		1:1	.021	.044	.003	.017	.033	.028	.015	.009
		3:1	.	.079	.	.009	.	.078	.	.008
	120	1:3	.002	.006	.001	.004	.000	.003	.001	.003
		1:1	.006	.010	.003	.006	.003	.005	.002	.004
		3:1	.	.026	.	.013	.	.015	.	.007

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.1.4.5

Tabelle C-16. MSE_{MC}

			$d_{s,R}^{S_2}$		$\hat{d}_R^{S_2}$		$d_{s,R}^{S_2,Boot}$		$\hat{d}_R^{S_2,Boot}$	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	40	1:3	.232	.306	.208	.275	.238	.315	.217	.288
		1:1	.249	.251	.206	.207	.245	.249	.214	.217
		3:1	.	.616	.	.388	.	113.309	.	110.523
	120	1:3	.066	.086	.064	.084	.067	.089	.066	.086
		1:1	.065	.062	.061	.058	.067	.064	.064	.061
		3:1	.	.089	.	.078	.	.095	.	.086
Gleich	40	1:3	.307	.387	.276	.350	.307	.401	.281	.368
		1:1	.323	.311	.269	.259	.289	.290	.254	.255
		3:1	.	.696	.	.441	.	44.766	.	31.940
	120	1:3	.092	.122	.089	.118	.093	.125	.090	.121
		1:1	.081	.079	.077	.075	.079	.079	.076	.076
		3:1	.	.101	.	.087	.	.098	.	.090
g-h-1	40	1:3	.152	.194	.136	.173	.162	.203	.147	.184
		1:1	.186	.198	.152	.163	.200	.217	.173	.188
		3:1	.	.559	.	.350	.	14.820	.	10.663
	120	1:3	.041	.053	.040	.051	.043	.055	.042	.054
		1:1	.046	.046	.043	.043	.050	.050	.048	.048
		3:1	.	.080	.	.070	.	.093	.	.084
g-h-2	40	1:3	.313	.308	.276	.273	.295	.285	.270	.260
		1:1	.434	.289	.357	.237	.410	.268	.359	.235
		3:1	.	.946	.	.609	.	42.944	.	30.620
	120	1:3	.079	.076	.076	.074	.079	.075	.076	.073
		1:1	.093	.065	.087	.061	.095	.065	.091	.062
		3:1	.	.113	.	.099	.	.116	.	.106
g-h-3	40	1:3	.165	.188	.145	.167	.173	.193	.157	.176
		1:1	.216	.190	.176	.154	.233	.203	.202	.175
		3:1	.	.567	.	.356	.	24.304	.	17.384
	120	1:3	.044	.049	.042	.047	.046	.051	.045	.049
		1:1	.052	.044	.049	.042	.057	.048	.054	.046
		3:1	.	.079	.	.069	.	.090	.	.082
g-h-4	40	1:3	.451	.451	.400	.403	.403	.414	.369	.379
		1:1	.621	.406	.514	.336	.531	.348	.467	.307
		3:1	.	1.748	.	1.165	.	68.580	.	88.813
	120	1:3	.107	.116	.103	.111	.104	.113	.101	.110
		1:1	.119	.086	.112	.081	.114	.082	.109	.079
		3:1	.	.145	.	.126	.	.136	.	.124
Hyper	40	1:3	.561	.654	.503	.586	.878	.935	.795	.847
		1:1	.619	.530	.521	.437	1.002	.735	.854	.621
		3:1	.	.569	.	.356	.	.872	.	.571
	120	1:3	.248	.211	.241	.204	.383	.258	.373	.251
		1:1	.289	.162	.274	.152	.475	.220	.454	.210
		3:1	.	.216	.	.187	.	.331	.	.296

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.1.4.5

Tabelle C-17. $|CP-.95|$

			NC (d_s)		Perc (d_s)		Perc (\hat{d})		BCA (d_s)		BCA (\hat{d})		Perc-t (d_s)		Perc-t (\hat{d})	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	40	1:3	.002	.208	.033	.128	.030	.122	.011	.047	.012	.043	.008	.053	.010	.050
		1:1	.005	.014	.019	.036	.017	.032	.006	.010	.006	.010	.004	.006	.004	.007
		3:1	.	.036	.	.031	.	.038	.	.048	.	.060	.	.067	.	.081
	120	1:3	.003	.274	.007	.130	.006	.126	.004	.069	.004	.066	.003	.078	.003	.075
		1:1	.006	.010	.012	.011	.012	.009	.005	.004	.005	.004	.006	.004	.006	.005
		3:1	.	.038	.	.085	.	.092	.	.116	.	.126	.	.134	.	.144
Gleich	40	1:3	.005	.177	.026	.082	.025	.077	.003	.019	.004	.017	.004	.013	.006	.012
		1:1	.007	.005	.011	.015	.010	.015	.007	.011	.006	.011	.016	.021	.013	.021
		3:1	.	.040	.	.043	.	.051	.	.048	.	.059	.	.076	.	.092
	120	1:3	.005	.263	.006	.117	.006	.113	.002	.062	.002	.058	.003	.052	.003	.049
		1:1	.007	.006	.005	.008	.004	.008	.003	.004	.002	.004	.004	.009	.003	.009
		3:1	.	.034	.	.112	.	.122	.	.128	.	.137	.	.154	.	.164
g-h-1	40	1:3	.036	.286	.080	.225	.073	.218	.039	.102	.036	.098	.047	.124	.047	.120
		1:1	.044	.072	.079	.107	.072	.100	.033	.045	.029	.041	.053	.065	.052	.063
		3:1	.	.031	.	.040	.	.037	.	.038	.	.041	.	.076	.	.082
	120	1:3	.048	.363	.044	.223	.041	.219	.026	.132	.025	.129	.038	.160	.037	.157
		1:1	.063	.093	.053	.067	.050	.064	.026	.036	.024	.035	.050	.059	.050	.059
		3:1	.	.053	.	.022	.	.023	.	.065	.	.069	.	.104	.	.110
g-h-2	40	1:3	.024	.149	.049	.166	.047	.156	.022	.051	.021	.046	.060	.029	.063	.025
		1:1	.031	.027	.063	.041	.058	.035	.025	.018	.024	.018	.041	.023	.042	.026
		3:1	.	.042	.	.024	.	.030	.	.069	.	.083	.	.089	.	.101
	120	1:3	.027	.254	.022	.240	.021	.233	.009	.154	.009	.148	.031	.137	.031	.132
		1:1	.033	.025	.029	.022	.026	.019	.013	.015	.013	.013	.023	.023	.023	.023
		3:1	.	.038	.	.099	.	.110	.	.175	.	.189	.	.185	.	.196
g-h-3	40	1:3	.051	.265	.093	.235	.087	.228	.046	.096	.043	.091	.069	.110	.068	.105
		1:1	.065	.047	.107	.110	.098	.101	.046	.044	.042	.041	.070	.053	.070	.052
		3:1	.	.032	.	.037	.	.034	.	.036	.	.039	.	.059	.	.066
	120	1:3	.078	.355	.063	.257	.060	.252	.039	.153	.037	.150	.064	.172	.063	.168
		1:1	.097	.074	.077	.084	.074	.079	.043	.044	.041	.042	.075	.066	.073	.064
		3:1	.	.046	.	.019	.	.020	.	.062	.	.067	.	.100	.	.104
g-h-4	40	1:3	.003	.132	.025	.108	.024	.100	.004	.024	.004	.019	.023	.009	.027	.011
		1:1	.003	.029	.025	.012	.022	.011	.007	.013	.006	.012	.004	.024	.005	.024
		3:1	.	.045	.	.053	.	.068	.	.081	.	.101	.	.090	.	.109
	120	1:3	.002	.235	.006	.195	.006	.187	.003	.142	.003	.135	.005	.113	.007	.107
		1:1	.003	.030	.003	.003	.003	.003	.004	.005	.004	.005	.004	.013	.005	.013
		3:1	.	.033	.	.182	.	.196	.	.214	.	.229	.	.209	.	.222
Hyper	40	1:3	.026	.163	.033	.064	.031	.059	.012	.035	.012	.033	.016	.035	.019	.034
		1:1	.027	.003	.028	.012	.026	.011	.009	.006	.009	.005	.013	.019	.015	.018
		3:1	.	.039	.	.043	.	.053	.	.050	.	.062	.	.069	.	.085
	120	1:3	.026	.258	.011	.108	.010	.105	.002	.063	.002	.060	.003	.042	.004	.038
		1:1	.028	.004	.008	.005	.007	.005	.006	.003	.006	.003	.009	.013	.010	.012
		3:1	.	.035	.	.116	.	.126	.	.132	.	.141	.	.152	.	.162

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.1.4.5

Tabelle C-18. Mittlere Intervalllänge

			NC (d_s)		Perc (d_s)		Perc (\hat{d})		BCA (d_s)		BCA (\hat{d})		Perc-t (d_s)		Perc-t (\hat{d})	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	40	1:3	1.508	1.572	1.536	2.564	1.505	2.513	1.520	2.515	1.490	2.465	1.478	2.318	1.448	2.272
		1:1	1.324	1.327	1.377	1.458	1.350	1.429	1.357	1.420	1.330	1.392	1.322	1.381	1.296	1.354
		3:1	.	1.484	.	1.067	.	1.046	.	1.046	.	1.025	.	1.041	.	1.020
	120	1:3	.865	.892	.871	1.388	.865	1.379	.868	1.380	.863	1.371	.861	1.349	.856	1.341
		1:1	.759	.759	.769	.798	.764	.793	.766	.793	.761	.787	.760	.786	.755	.781
		3:1	.	.854	.	.597	.	.594	.	.594	.	.590	.	.593	.	.589
Gleich	40	1:3	1.506	1.566	1.513	2.653	1.483	2.600	1.500	2.564	1.470	2.513	1.461	2.388	1.432	2.340
		1:1	1.322	1.323	1.356	1.431	1.329	1.402	1.337	1.392	1.310	1.364	1.306	1.361	1.280	1.334
		3:1	.	1.483	.	1.030	.	1.010	.	1.014	.	.994	.	1.008	.	.988
	120	1:3	.864	.890	.853	1.370	.848	1.361	.851	1.356	.846	1.347	.845	1.333	.839	1.325
		1:1	.758	.759	.749	.769	.745	.764	.747	.763	.742	.759	.741	.759	.737	.754
		3:1	.	.853	.	.572	.	.569	.	.570	.	.566	.	.569	.	.565
g-h-1	40	1:3	1.522	1.604	1.632	2.581	1.600	2.530	1.598	2.525	1.566	2.474	1.540	2.288	1.510	2.242
		1:1	1.340	1.347	1.492	1.589	1.462	1.558	1.458	1.533	1.429	1.502	1.399	1.463	1.371	1.434
		3:1	.	1.497	.	1.212	.	1.188	.	1.167	.	1.144	.	1.157	.	1.134
	120	1:3	.869	.901	.957	1.468	.951	1.459	.955	1.466	.949	1.456	.937	1.405	.931	1.396
		1:1	.763	.765	.869	.916	.864	.910	.868	.910	.863	.904	.849	.888	.843	.883
		3:1	.	.857	.	.708	.	.703	.	.702	.	.698	.	.696	.	.691
g-h-2	40	1:3	1.516	1.569	1.674	2.164	1.641	2.121	1.614	2.110	1.582	2.068	1.571	1.991	1.540	1.951
		1:1	1.333	1.328	1.487	1.231	1.457	1.206	1.448	1.205	1.420	1.181	1.395	1.173	1.368	1.150
		3:1	.	1.485	.	.929	.	.910	.	.921	.	.903	.	.906	.	.888
	120	1:3	.866	.891	.944	1.122	.938	1.115	.934	1.116	.928	1.109	.926	1.097	.920	1.090
		1:1	.761	.760	.842	.653	.837	.649	.838	.646	.833	.642	.826	.644	.820	.640
		3:1	.	.854	.	.510	.	.507	.	.508	.	.505	.	.506	.	.503
g-h-3	40	1:3	1.530	1.603	1.693	2.456	1.660	2.407	1.638	2.397	1.606	2.349	1.584	2.194	1.552	2.150
		1:1	1.348	1.348	1.531	1.492	1.501	1.462	1.492	1.447	1.462	1.418	1.425	1.382	1.396	1.354
		3:1	.	1.498	.	1.139	.	1.116	.	1.110	.	1.088	.	1.090	.	1.069
	120	1:3	.871	.902	.995	1.361	.988	1.352	.987	1.364	.981	1.355	.969	1.307	.963	1.299
		1:1	.766	.767	.903	.842	.898	.837	.901	.842	.896	.837	.877	.818	.872	.812
		3:1	.	.858	.	.654	.	.650	.	.654	.	.650	.	.643	.	.639
g-h-4	40	1:3	1.508	1.558	1.583	2.216	1.552	2.172	1.555	2.159	1.524	2.116	1.510	2.049	1.480	2.008
		1:1	1.325	1.320	1.393	1.200	1.366	1.176	1.370	1.169	1.343	1.146	1.334	1.157	1.308	1.134
		3:1	.	1.481	.	.883	.	.865	.	.872	.	.855	.	.869	.	.852
	120	1:3	.865	.889	.878	1.142	.872	1.135	.875	1.132	.869	1.125	.866	1.120	.861	1.113
		1:1	.759	.758	.772	.635	.767	.631	.768	.631	.764	.627	.762	.629	.757	.625
		3:1	.	.853	.	.485	.	.482	.	.484	.	.481	.	.483	.	.480
Hyper	40	1:3	1.513	1.566	1.735	2.510	1.701	2.460	1.688	2.397	1.655	2.349	1.635	2.276	1.602	2.231
		1:1	1.326	1.323	1.513	1.418	1.483	1.390	1.483	1.382	1.454	1.354	1.438	1.350	1.410	1.323
		3:1	.	1.482	.	1.021	.	1.001	.	1.009	.	.989	.	1.000	.	.980
	120	1:3	.866	.890	.971	1.379	.965	1.370	.961	1.364	.955	1.356	.954	1.341	.948	1.333
		1:1	.759	.759	.854	.768	.848	.763	.847	.763	.842	.759	.840	.758	.834	.753
		3:1	.	.853	.	.569	.	.566	.	.567	.	.564	.	.566	.	.562

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.1.4.5

Tabelle C-19. $|CP - .95|$

			NC (d'_s)		Perc (d'_s)		Perc (\hat{d}')		BCA (d'_s)		BCA (\hat{d}')		Perc-t (d'_s)		Perc-t (\hat{d}')	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	40	1:3	.005	.004	.041	.076	.034	.072	.009	.022	.010	.021	.006	.008	.007	.010
		1:1	.006	.004	.019	.036	.017	.031	.005	.008	.006	.009	.004	.005	.004	.006
		3:1	.	.001	.	.022	.	.017	.	.005	.	.013	.	.003	.	.016
	120	1:3	.008	.004	.011	.023	.009	.022	.004	.007	.004	.007	.003	.004	.003	.004
		1:1	.005	.003	.012	.011	.012	.009	.005	.003	.005	.004	.006	.003	.007	.005
		3:1	.	.002	.	.008	.	.006	.	.003	.	.004	.	.002	.	.003
Gleich	40	1:3	.007	.009	.028	.039	.026	.038	.007	.009	.007	.009	.019	.025	.017	.024
		1:1	.007	.009	.011	.015	.010	.015	.007	.009	.006	.009	.014	.019	.011	.019
		3:1	.	.009	.	.018	.	.017	.	.003	.	.010	.	.009	.	.017
	120	1:3	.011	.010	.006	.011	.006	.010	.004	.007	.004	.006	.010	.013	.009	.013
		1:1	.007	.010	.005	.008	.004	.007	.003	.004	.002	.004	.004	.008	.003	.008
		3:1	.	.006	.	.006	.	.005	.	.001	.	.003	.	.004	.	.006
g-h-1	40	1:3	.028	.034	.097	.156	.083	.148	.041	.064	.039	.061	.059	.066	.058	.066
		1:1	.041	.053	.079	.107	.071	.096	.033	.048	.030	.042	.057	.068	.056	.066
		3:1	.	.055	.	.091	.	.065	.	.042	.	.034	.	.070	.	.068
	120	1:3	.043	.058	.057	.092	.052	.089	.030	.049	.028	.048	.054	.074	.053	.073
		1:1	.062	.077	.053	.067	.049	.061	.026	.037	.024	.035	.051	.062	.050	.061
		3:1	.	.076	.	.056	.	.045	.	.032	.	.028	.	.055	.	.053
g-h-2	40	1:3	.004	.031	.049	.062	.041	.054	.012	.022	.014	.023	.020	.024	.025	.026
		1:1	.028	.031	.063	.041	.057	.031	.024	.018	.023	.019	.051	.022	.051	.029
		3:1	.	.024	.	.053	.	.023	.	.024	.	.027	.	.026	.	.039
	120	1:3	.008	.034	.022	.033	.020	.030	.009	.021	.008	.019	.010	.030	.010	.031
		1:1	.031	.028	.029	.022	.026	.018	.012	.014	.013	.013	.026	.022	.026	.024
		3:1	.	.021	.	.026	.	.017	.	.015	.	.015	.	.018	.	.019
g-h-3	40	1:3	.033	.014	.106	.152	.091	.143	.045	.057	.043	.055	.062	.049	.064	.050
		1:1	.061	.032	.107	.110	.096	.095	.046	.045	.043	.040	.075	.057	.074	.056
		3:1	.	.037	.	.104	.	.071	.	.046	.	.035	.	.066	.	.060
	120	1:3	.049	.032	.069	.100	.064	.097	.038	.052	.036	.051	.062	.067	.061	.066
		1:1	.095	.059	.077	.084	.074	.076	.043	.045	.041	.042	.076	.069	.074	.067
		3:1	.	.071	.	.077	.	.063	.	.044	.	.038	.	.071	.	.067
g-h-4	40	1:3	.015	.030	.023	.030	.021	.028	.009	.013	.007	.013	.018	.023	.015	.026
		1:1	.001	.030	.025	.012	.022	.011	.006	.011	.006	.011	.011	.022	.012	.023
		3:1	.	.027	.	.020	.	.015	.	.008	.	.017	.	.006	.	.019
	120	1:3	.021	.032	.004	.007	.004	.008	.005	.007	.005	.007	.011	.018	.010	.019
		1:1	.004	.032	.003	.003	.003	.003	.005	.005	.004	.005	.004	.013	.005	.012
		3:1	.	.027	.	.009	.	.008	.	.004	.	.009	.	.004	.	.005
Hyper	40	1:3	.008	.014	.034	.026	.031	.026	.008	.031	.010	.032	.005	.036	.006	.035
		1:1	.025	.009	.028	.012	.025	.011	.009	.006	.009	.005	.015	.021	.016	.018
		3:1	.	.008	.	.011	.	.012	.	.004	.	.010	.	.013	.	.018
	120	1:3	.005	.013	.011	.007	.010	.007	.002	.006	.001	.006	.003	.015	.002	.015
		1:1	.027	.009	.008	.005	.007	.005	.006	.002	.006	.002	.010	.011	.011	.011
		3:1	.	.005	.	.005	.	.006	.	.003	.	.004	.	.006	.	.008

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.1.4.5

Tabelle C-20. Mittlere Intervalllänge

			NC (d'_s)		Perc (d'_s)		Perc (\hat{d}')		BCA (d'_s)		BCA (\hat{d}')		Perc-t (d'_s)		Perc-t (\hat{d}')	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	40	1:3	1.583	1.942	1.582	2.169	1.522	2.119	1.556	2.055	1.498	2.010	1.576	1.918	1.525	1.875
		1:1	1.328	1.374	1.377	1.458	1.347	1.415	1.357	1.418	1.328	1.375	1.315	1.373	1.288	1.331
		3:1	.	1.207	.	1.256	.	1.173	.	1.233	.	1.151	.	1.200	.	1.120
	120	1:3	.908	1.095	.885	1.118	.876	1.110	.881	1.100	.872	1.093	.881	1.076	.873	1.069
		1:1	.760	.784	.769	.798	.764	.790	.766	.792	.761	.784	.759	.784	.754	.776
		3:1	.	.689	.	.702	.	.688	.	.698	.	.684	.	.693	.	.679
Gleich	40	1:3	1.590	1.933	1.587	2.242	1.525	2.192	1.557	2.092	1.499	2.047	1.593	1.955	1.542	1.912
		1:1	1.324	1.370	1.356	1.431	1.328	1.388	1.337	1.393	1.309	1.350	1.300	1.351	1.274	1.309
		3:1	.	1.202	.	1.221	.	1.138	.	1.202	.	1.121	.	1.174	.	1.094
	120	1:3	.910	1.094	.869	1.088	.860	1.080	.864	1.071	.856	1.064	.865	1.049	.857	1.042
		1:1	.759	.784	.749	.769	.744	.761	.747	.764	.742	.756	.741	.756	.736	.748
		3:1	.	.688	.	.676	.	.662	.	.673	.	.659	.	.669	.	.655
g-h-1	40	1:3	1.580	1.982	1.651	2.236	1.587	2.181	1.611	2.106	1.548	2.056	1.603	1.967	1.549	1.921
		1:1	1.351	1.399	1.492	1.589	1.458	1.545	1.457	1.529	1.424	1.485	1.391	1.455	1.361	1.412
		3:1	.	1.240	.	1.399	.	1.312	.	1.356	.	1.269	.	1.295	.	1.212
	120	1:3	.905	1.107	.963	1.223	.953	1.215	.961	1.200	.951	1.192	.954	1.157	.944	1.149
		1:1	.766	.791	.869	.916	.863	.908	.868	.909	.862	.901	.847	.885	.842	.877
		3:1	.	.698	.	.820	.	.804	.	.817	.	.801	.	.795	.	.779
g-h-2	40	1:3	1.573	1.915	1.575	1.803	1.525	1.757	1.524	1.704	1.476	1.664	1.542	1.709	1.508	1.672
		1:1	1.343	1.375	1.487	1.231	1.451	1.192	1.450	1.204	1.414	1.167	1.375	1.164	1.345	1.125
		3:1	.	1.220	.	1.107	.	1.026	.	1.100	.	1.018	.	1.045	.	.968
	120	1:3	.904	1.091	.861	.885	.855	.878	.854	.875	.848	.869	.865	.868	.860	.862
		1:1	.763	.785	.842	.653	.836	.646	.838	.647	.832	.640	.822	.642	.816	.635
		3:1	.	.692	.	.616	.	.601	.	.616	.	.601	.	.603	.	.588
g-h-3	40	1:3	1.591	1.971	1.677	2.105	1.616	2.053	1.621	1.979	1.562	1.933	1.611	1.879	1.561	1.837
		1:1	1.360	1.402	1.531	1.492	1.497	1.449	1.491	1.443	1.457	1.401	1.416	1.372	1.385	1.331
		3:1	.	1.246	.	1.327	.	1.241	.	1.298	.	1.212	.	1.235	.	1.153
	120	1:3	.907	1.107	.968	1.117	.959	1.110	.965	1.106	.956	1.099	.954	1.065	.946	1.058
		1:1	.770	.794	.903	.842	.897	.834	.901	.842	.895	.834	.876	.814	.870	.806
		3:1	.	.702	.	.768	.	.752	.	.770	.	.754	.	.744	.	.728
g-h-4	40	1:3	1.582	1.908	1.507	1.829	1.461	1.781	1.474	1.729	1.432	1.689	1.511	1.732	1.481	1.694
		1:1	1.330	1.366	1.393	1.200	1.360	1.161	1.371	1.169	1.338	1.130	1.315	1.148	1.287	1.109
		3:1	.	1.204	.	1.062	.	.982	.	1.053	.	.972	.	1.019	.	.942
	120	1:3	.908	1.089	.816	.880	.810	.874	.812	.865	.807	.859	.817	.863	.812	.857
		1:1	.759	.783	.772	.635	.766	.628	.768	.631	.763	.624	.760	.627	.755	.620
		3:1	.	.688	.	.587	.	.573	.	.587	.	.573	.	.581	.	.567
Hyper	40	1:3	1.599	1.934	1.708	2.088	1.652	2.039	1.660	1.957	1.608	1.911	1.658	1.853	1.613	1.811
		1:1	1.332	1.371	1.513	1.418	1.479	1.374	1.484	1.384	1.450	1.340	1.423	1.338	1.393	1.295
		3:1	.	1.201	.	1.217	.	1.131	.	1.202	.	1.118	.	1.170	.	1.088
	120	1:3	.912	1.093	.944	1.085	.936	1.077	.935	1.070	.927	1.062	.932	1.045	.924	1.038
		1:1	.761	.784	.854	.768	.848	.760	.847	.764	.841	.756	.837	.755	.831	.747
		3:1	.	.688	.	.679	.	.664	.	.677	.	.662	.	.672	.	.657

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.1.4.5

Tabelle C-21. $|CP - .95|$

			NC ($d_s^{S_1}$)		Perc ($d_s^{S_1}$)		Perc (\hat{d}^{S_1})		BCA ($d_s^{S_1}$)		BCA (\hat{d}^{S_1})		Perc-t ($d_s^{S_1}$)		Perc-t (\hat{d}^{S_1})	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	40	1:3	.006	.003	.064	.080	.048	.066	.016	.021	.019	.023	.016	.007	.025	.009
		1:1	.004	.003	.023	.038	.018	.032	.008	.010	.009	.012	.008	.004	.013	.007
		3:1	.	.002	.	.022	.	.018	.	.005	.	.007	.	.006	.	.009
	120	1:3	.005	.002	.019	.024	.015	.020	.009	.008	.010	.009	.008	.004	.010	.006
		1:1	.004	.003	.013	.011	.011	.010	.004	.004	.006	.005	.006	.004	.008	.005
		3:1	.	.001	.	.008	.	.007	.	.002	.	.003	.	.003	.	.002
Gleich	40	1:3	.016	.010	.034	.040	.027	.036	.016	.011	.021	.013	.028	.029	.040	.031
		1:1	.013	.011	.011	.016	.009	.015	.010	.010	.007	.010	.016	.021	.015	.021
		3:1	.	.009	.	.017	.	.015	.	.003	.	.004	.	.010	.	.011
	120	1:3	.015	.011	.008	.012	.006	.010	.007	.007	.007	.006	.014	.015	.016	.015
		1:1	.011	.012	.005	.008	.005	.008	.004	.003	.004	.004	.005	.009	.005	.008
		3:1	.	.009	.	.006	.	.005	.	.001	.	.001	.	.004	.	.003
g-h-1	40	1:3	.038	.040	.150	.171	.120	.144	.053	.059	.047	.059	.031	.036	.041	.039
		1:1	.047	.053	.087	.115	.072	.100	.042	.053	.037	.046	.063	.061	.063	.061
		3:1	.	.057	.	.090	.	.080	.	.049	.	.045	.	.076	.	.074
	120	1:3	.078	.068	.099	.103	.085	.090	.053	.052	.047	.048	.057	.065	.058	.065
		1:1	.080	.084	.062	.073	.056	.067	.038	.042	.036	.039	.061	.061	.060	.060
		3:1	.	.079	.	.056	.	.052	.	.035	.	.033	.	.059	.	.058
g-h-2	40	1:3	.029	.034	.081	.081	.049	.050	.026	.030	.028	.030	.032	.025	.047	.044
		1:1	.020	.034	.057	.049	.044	.033	.019	.023	.019	.021	.028	.029	.033	.039
		3:1	.	.029	.	.046	.	.034	.	.021	.	.020	.	.030	.	.033
	120	1:3	.021	.034	.050	.047	.037	.032	.028	.026	.026	.024	.037	.034	.043	.040
		1:1	.018	.030	.029	.029	.022	.022	.014	.018	.013	.017	.024	.027	.026	.030
		3:1	.	.027	.	.023	.	.019	.	.014	.	.013	.	.019	.	.019
g-h-3	40	1:3	.030	.017	.164	.175	.129	.141	.057	.056	.053	.055	.027	.022	.034	.027
		1:1	.040	.030	.104	.121	.089	.101	.047	.051	.043	.044	.058	.048	.057	.048
		3:1	.	.034	.	.103	.	.090	.	.052	.	.046	.	.065	.	.062
	120	1:3	.060	.043	.116	.119	.100	.103	.057	.059	.050	.053	.060	.063	.058	.059
		1:1	.072	.060	.084	.092	.076	.083	.049	.051	.046	.046	.068	.069	.066	.066
		3:1	.	.068	.	.079	.	.073	.	.046	.	.044	.	.072	.	.069
g-h-4	40	1:3	.034	.032	.027	.032	.011	.026	.020	.014	.018	.015	.028	.034	.046	.045
		1:1	.030	.032	.018	.010	.011	.012	.009	.014	.009	.011	.011	.021	.011	.026
		3:1	.	.032	.	.015	.	.012	.	.006	.	.006	.	.011	.	.011
	120	1:3	.034	.033	.005	.007	.003	.009	.009	.008	.007	.009	.016	.020	.019	.025
		1:1	.031	.033	.001	.002	.001	.003	.005	.005	.004	.006	.009	.013	.008	.014
		3:1	.	.031	.	.007	.	.006	.	.004	.	.004	.	.003	.	.004
Hyper	40	1:3	.011	.016	.048	.027	.038	.026	.016	.032	.020	.034	.019	.033	.029	.041
		1:1	.010	.012	.021	.011	.018	.011	.004	.006	.005	.005	.007	.020	.009	.018
		3:1	.	.014	.	.010	.	.009	.	.003	.	.003	.	.014	.	.014
	120	1:3	.009	.014	.016	.007	.013	.008	.006	.006	.007	.006	.010	.017	.011	.017
		1:1	.007	.013	.007	.005	.006	.006	.005	.002	.005	.002	.007	.012	.007	.012
		3:1	.	.011	.	.004	.	.004	.	.001	.	.002	.	.006	.	.006

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.1.4.5

Tabelle C-22. Mittlere Intervalllänge

			NC($d_s^{S_1}$)		Perc($d_s^{S_1}$)		Perc(\hat{d}^{S_1})		BCA($d_s^{S_1}$)		BCA(\hat{d}^{S_1})		Perc-t($d_s^{S_1}$)		Perc-t(\hat{d}^{S_1})	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	40	1:3	1.785	1.995	2.260	2.497	2.066	2.283	1.985	2.234	1.810	2.037	1.837	2.023	1.679	1.849
		1:1	1.427	1.413	1.544	1.531	1.482	1.469	1.485	1.469	1.426	1.410	1.405	1.412	1.349	1.355
		3:1	.	1.234	.	1.285	.	1.252	.	1.257	.	1.225	.	1.220	.	1.188
	120	1:3	.977	1.108	1.027	1.164	1.000	1.133	.995	1.134	.969	1.105	.978	1.104	.953	1.075
		1:1	.803	.800	.822	.818	.811	.807	.813	.809	.802	.799	.800	.799	.790	.789
		3:1	.	.704	.	.713	.	.707	.	.709	.	.703	.	.703	.	.697
Gleich	40	1:3	1.745	1.968	2.188	2.523	2.001	2.307	1.923	2.237	1.760	2.047	1.802	2.022	1.648	1.849
		1:1	1.408	1.400	1.455	1.476	1.397	1.417	1.415	1.426	1.358	1.369	1.355	1.373	1.301	1.318
		3:1	.	1.227	.	1.233	.	1.201	.	1.214	.	1.183	.	1.185	.	1.154
	120	1:3	.969	1.104	.958	1.115	.933	1.086	.937	1.092	.913	1.064	.921	1.063	.897	1.036
		1:1	.799	.798	.775	.779	.765	.769	.770	.773	.760	.763	.761	.764	.751	.754
		3:1	.	.702	.	.681	.	.675	.	.678	.	.672	.	.674	.	.668
g-h-1	40	1:3	1.970	2.118	2.767	2.849	2.530	2.605	2.339	2.446	2.129	2.220	2.067	2.206	1.890	2.017
		1:1	1.539	1.472	1.889	1.758	1.813	1.688	1.777	1.647	1.706	1.581	1.617	1.544	1.553	1.483
		3:1	.	1.283	.	1.494	.	1.455	.	1.433	.	1.395	.	1.355	.	1.320
	120	1:3	1.029	1.140	1.269	1.346	1.236	1.310	1.202	1.290	1.169	1.256	1.152	1.229	1.122	1.197
		1:1	.836	.818	1.016	.974	1.004	.962	.992	.956	.979	.943	.946	.926	.934	.914
		3:1	.	.717	.	.858	.	.851	.	.849	.	.842	.	.821	.	.814
g-h-2	40	1:3	1.894	2.020	2.293	2.266	2.097	2.071	2.003	1.952	1.822	1.776	1.697	1.812	1.552	1.656
		1:1	1.498	1.431	1.547	1.306	1.485	1.254	1.538	1.259	1.476	1.207	1.336	1.172	1.283	1.125
		3:1	.	1.254	.	1.118	.	1.089	.	1.111	.	1.082	.	1.034	.	1.007
	120	1:3	.998	1.114	.928	.938	.904	.913	.908	.921	.883	.895	.848	.884	.826	.861
		1:1	.817	.804	.775	.660	.765	.652	.773	.651	.763	.643	.732	.638	.723	.630
		3:1	.	.708	.	.602	.	.596	.	.600	.	.595	.	.584	.	.579
g-h-3	40	1:3	2.013	2.117	2.683	2.661	2.453	2.433	2.303	2.291	2.093	2.079	2.027	2.093	1.853	1.914
		1:1	1.569	1.484	1.839	1.635	1.766	1.569	1.764	1.551	1.693	1.487	1.587	1.449	1.524	1.391
		3:1	.	1.296	.	1.395	.	1.358	.	1.359	.	1.324	.	1.278	.	1.245
	120	1:3	1.044	1.147	1.202	1.223	1.170	1.191	1.153	1.188	1.122	1.156	1.087	1.125	1.059	1.096
		1:1	.848	.823	.974	.883	.962	.872	.964	.875	.951	.864	.908	.840	.896	.829
		3:1	.	.723	.	.786	.	.780	.	.787	.	.780	.	.756	.	.750
g-h-4	40	1:3	1.778	1.967	2.122	2.273	1.940	2.078	1.810	1.930	1.650	1.765	1.569	1.788	1.435	1.635
		1:1	1.428	1.400	1.358	1.239	1.304	1.189	1.333	1.187	1.280	1.139	1.221	1.144	1.172	1.098
		3:1	.	1.229	.	1.051	.	1.023	.	1.038	.	1.011	.	1.000	.	.974
	120	1:3	.973	1.102	.814	.895	.792	.872	.791	.872	.770	.849	.772	.861	.752	.838
		1:1	.802	.798	.690	.628	.681	.620	.689	.623	.680	.615	.673	.617	.664	.609
		3:1	.	.702	.	.570	.	.565	.	.569	.	.564	.	.562	.	.558
Hyper	40	1:3	1.763	1.962	2.109	2.381	1.928	2.177	1.926	2.180	1.760	1.997	1.768	1.911	1.616	1.747
		1:1	1.419	1.398	1.489	1.449	1.429	1.391	1.455	1.408	1.397	1.352	1.371	1.349	1.316	1.295
		3:1	.	1.225	.	1.219	.	1.187	.	1.205	.	1.173	.	1.170	.	1.139
	120	1:3	.974	1.102	.994	1.105	.968	1.076	.968	1.086	.943	1.058	.949	1.054	.924	1.027
		1:1	.801	.798	.796	.772	.786	.762	.791	.767	.781	.757	.778	.757	.768	.747
		3:1	.	.701	.	.674	.	.668	.	.672	.	.666	.	.667	.	.661

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.1.4.5

Tabelle C-23. $|CP - .95|$

			NC ($d_s^{S_2}$)		Perc ($d_s^{S_2}$)		Perc (\hat{d}^{S_2})		BCA ($d_s^{S_2}$)		BCA (\hat{d}^{S_2})		Perc-t ($d_s^{S_2}$)		Perc-t (\hat{d}^{S_2})	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	40	1:3	.007	.019	.029	.045	.027	.043	.014	.032	.015	.032	.015	.037	.016	.035
		1:1	.004	.004	.022	.024	.017	.021	.008	.004	.009	.007	.008	.011	.011	.014
		3:1	.	.008	.	.043	.	.027	.	.016	.	.021	.	.022	.	.044
	120	1:3	.003	.004	.006	.011	.005	.011	.003	.006	.003	.007	.003	.009	.003	.009
		1:1	.003	.002	.013	.007	.011	.006	.006	.002	.007	.002	.008	.003	.010	.004
		3:1	.	.004	.	.014	.	.010	.	.009	.	.010	.	.010	.	.016
Gleich	40	1:3	.005	.012	.025	.029	.024	.028	.004	.005	.004	.005	.007	.016	.007	.015
		1:1	.011	.007	.012	.015	.010	.013	.008	.007	.007	.004	.013	.008	.014	.005
		3:1	.	.020	.	.021	.	.012	.	.016	.	.017	.	.017	.	.035
	120	1:3	.005	.004	.006	.009	.005	.009	.002	.002	.002	.002	.004	.005	.004	.005
		1:1	.011	.009	.004	.005	.004	.005	.003	.002	.002	.003	.005	.004	.003	.003
		3:1	.	.018	.	.004	.	.003	.	.006	.	.005	.	.006	.	.012
g-h-1	40	1:3	.046	.045	.067	.064	.060	.059	.048	.059	.045	.056	.073	.071	.072	.068
		1:1	.048	.049	.086	.077	.074	.066	.042	.048	.037	.043	.063	.066	.064	.066
		3:1	.	.048	.	.132	.	.101	.	.050	.	.044	.	.065	.	.075
	120	1:3	.058	.054	.036	.035	.033	.033	.034	.040	.033	.039	.052	.048	.051	.047
		1:1	.078	.079	.061	.049	.055	.043	.038	.038	.035	.035	.059	.056	.058	.054
		3:1	.	.098	.	.098	.	.084	.	.055	.	.049	.	.072	.	.071
g-h-2	40	1:3	.081	.080	.048	.052	.045	.050	.039	.039	.037	.040	.075	.042	.076	.042
		1:1	.097	.076	.071	.057	.062	.050	.039	.030	.038	.030	.079	.050	.083	.053
		3:1	.	.094	.	.128	.	.104	.	.068	.	.063	.	.090	.	.101
	120	1:3	.076	.056	.021	.022	.021	.021	.019	.018	.019	.017	.032	.015	.033	.015
		1:1	.108	.080	.036	.029	.032	.025	.022	.021	.021	.020	.039	.025	.041	.025
		3:1	.	.122	.	.068	.	.060	.	.041	.	.038	.	.057	.	.061
g-h-3	40	1:3	.081	.073	.084	.075	.076	.070	.064	.068	.059	.066	.103	.086	.099	.084
		1:1	.098	.081	.122	.095	.108	.083	.071	.063	.064	.056	.103	.088	.100	.086
		3:1	.	.079	.	.170	.	.135	.	.078	.	.066	.	.090	.	.090
	120	1:3	.112	.080	.060	.042	.057	.039	.054	.047	.053	.047	.083	.059	.082	.058
		1:1	.152	.127	.094	.076	.087	.070	.063	.060	.059	.057	.097	.082	.095	.079
		3:1	.	.153	.	.132	.	.117	.	.082	.	.074	.	.102	.	.099
g-h-4	40	1:3	.039	.053	.028	.045	.027	.047	.009	.020	.011	.021	.023	.017	.024	.016
		1:1	.038	.026	.030	.020	.026	.018	.009	.009	.012	.010	.020	.003	.028	.006
		3:1	.	.040	.	.063	.	.049	.	.029	.	.033	.	.037	.	.060
	120	1:3	.025	.019	.006	.012	.006	.012	.002	.002	.001	.002	.004	.006	.004	.006
		1:1	.036	.017	.005	.006	.005	.006	.006	.005	.005	.004	.005	.004	.008	.003
		3:1	.	.033	.	.019	.	.017	.	.010	.	.011	.	.015	.	.023
Hyper	40	1:3	.043	.020	.032	.022	.031	.022	.015	.030	.015	.030	.021	.045	.022	.044
		1:1	.050	.013	.038	.019	.035	.015	.019	.009	.019	.009	.032	.019	.037	.019
		3:1	.	.023	.	.047	.	.030	.	.029	.	.030	.	.039	.	.054
	120	1:3	.038	.011	.010	.007	.010	.007	.002	.005	.003	.005	.005	.008	.005	.007
		1:1	.054	.015	.011	.010	.010	.010	.006	.006	.007	.006	.014	.006	.015	.007
		3:1	.	.024	.	.025	.	.019	.	.015	.	.016	.	.020	.	.026

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.1.4.5

Tabelle C-24. Mittlere Intervalllänge

			NC ($d_s^{S_2}$)		Perc ($d_s^{S_2}$)		Perc (\hat{d}^{S_2})		BCA ($d_s^{S_2}$)		BCA (\hat{d}^{S_2})		Perc-t ($d_s^{S_2}$)		Perc-t (\hat{d}^{S_2})	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	40	1:3	1.541	1.787	1.577	1.810	1.536	1.763	1.562	1.810	1.521	1.763	1.627	1.927	1.584	1.877
		1:1	1.428	1.435	1.543	1.552	1.481	1.490	1.485	1.498	1.426	1.438	1.405	1.445	1.348	1.387
		3:1	.	1.528	.	1.965	.	1.797	.	1.683	.	1.537	.	1.461	.	1.336
	120	1:3	.879	1.021	.887	1.028	.880	1.020	.884	1.028	.877	1.019	.891	1.039	.883	1.030
		1:1	.804	.805	.823	.823	.813	.813	.814	.815	.804	.805	.801	.806	.791	.796
		3:1	.	.822	.	.869	.	.846	.	.834	.	.812	.	.807	.	.785
Gleich	40	1:3	1.530	1.780	1.501	1.724	1.462	1.679	1.494	1.729	1.455	1.684	1.587	1.882	1.546	1.833
		1:1	1.408	1.413	1.454	1.423	1.396	1.366	1.414	1.394	1.358	1.338	1.355	1.373	1.301	1.318
		3:1	.	1.473	.	1.773	.	1.621	.	1.536	.	1.405	.	1.370	.	1.253
	120	1:3	.877	1.021	.853	.995	.846	.987	.852	.995	.845	.987	.864	1.016	.857	1.007
		1:1	.800	.800	.775	.770	.766	.760	.771	.767	.761	.757	.762	.764	.752	.754
		3:1	.	.811	.	.772	.	.752	.	.754	.	.734	.	.736	.	.717
g-h-1	40	1:3	1.591	1.858	1.832	2.128	1.784	2.072	1.800	2.142	1.753	2.086	1.813	2.270	1.766	2.211
		1:1	1.532	1.571	1.883	1.973	1.808	1.894	1.766	1.876	1.695	1.801	1.606	1.735	1.541	1.665
		3:1	.	1.781	.	2.668	.	2.439	.	2.221	.	2.024	.	1.818	.	1.662
	120	1:3	.898	1.036	1.043	1.189	1.035	1.179	1.040	1.196	1.031	1.185	1.024	1.186	1.015	1.176
		1:1	.834	.842	1.014	1.042	1.001	1.029	.990	1.019	.977	1.006	.945	.978	.933	.965
		3:1	.	.893	.	1.177	.	1.146	.	1.101	.	1.072	.	1.010	.	.984
g-h-2	40	1:3	1.580	1.820	2.087	2.182	2.032	2.125	1.988	2.195	1.936	2.137	2.076	2.660	2.022	2.591
		1:1	1.518	1.530	2.364	2.154	2.270	2.067	2.078	1.981	1.995	1.901	1.835	1.911	1.761	1.835
		3:1	.	1.782	.	3.301	.	3.018	.	2.560	.	2.344	.	1.983	.	1.813
	120	1:3	.888	1.026	1.131	1.194	1.122	1.183	1.110	1.203	1.101	1.193	1.124	1.270	1.115	1.259
		1:1	.822	.826	1.184	1.089	1.168	1.075	1.121	1.049	1.106	1.035	1.076	1.037	1.062	1.024
		3:1	.	.870	.	1.332	.	1.297	.	1.200	.	1.169	.	1.102	.	1.073
g-h-3	40	1:3	1.633	1.871	2.032	2.227	1.979	2.169	1.968	2.227	1.916	2.169	2.063	2.639	2.010	2.570
		1:1	1.585	1.616	2.140	2.150	2.054	2.064	1.967	2.022	1.888	1.941	1.794	1.947	1.722	1.869
		3:1	.	1.866	.	2.956	.	2.703	.	2.420	.	2.208	.	2.014	.	1.841
	120	1:3	.907	1.047	1.167	1.273	1.157	1.263	1.150	1.279	1.140	1.268	1.149	1.312	1.139	1.301
		1:1	.851	.859	1.176	1.152	1.161	1.137	1.129	1.116	1.115	1.102	1.074	1.086	1.060	1.072
		3:1	.	.924	.	1.329	.	1.294	.	1.228	.	1.196	.	1.123	.	1.094
g-h-4	40	1:3	1.540	1.776	1.799	1.903	1.752	1.853	1.759	1.937	1.713	1.886	1.844	2.335	1.796	2.274
		1:1	1.442	1.436	2.000	1.792	1.920	1.721	1.803	1.681	1.732	1.614	1.640	1.631	1.574	1.566
		3:1	.	1.587	.	3.024	.	2.765	.	2.244	.	2.058	.	1.713	.	1.566
	120	1:3	.878	1.022	.971	1.063	.963	1.054	.968	1.073	.960	1.064	.979	1.116	.970	1.107
		1:1	.804	.805	.968	.897	.955	.886	.944	.886	.932	.874	.922	.879	.911	.868
		3:1	.	.826	.	1.051	.	1.024	.	.977	.	.952	.	.932	.	.907
Hyper	40	1:3	1.561	1.821	1.929	1.768	1.879	1.721	1.846	1.729	1.798	1.684	1.821	1.858	1.774	1.809
		1:1	1.453	1.450	2.084	1.598	2.001	1.534	1.926	1.541	1.850	1.480	1.677	1.482	1.610	1.422
		3:1	.	1.529	.	1.801	.	1.647	.	1.663	.	1.522	.	1.456	.	1.331
	120	1:3	.882	1.028	1.037	1.057	1.029	1.048	1.019	1.053	1.010	1.044	1.013	1.068	1.005	1.059
		1:1	.807	.807	1.054	.869	1.041	.858	1.014	.857	1.002	.846	.981	.847	.969	.836
		3:1	.	.825	.	.938	.	.914	.	.897	.	.874	.	.862	.	.839

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.1.4.5

Tabelle C-25. $|CP - .95|$

			NC($d_{s,R}$)		Perc($d_{s,R}$)		Perc(\hat{d}_R)		BCA($d_{s,R}$)		BCA(\hat{d}_R)		Perc-t($d_{s,R}$)		Perc-t(\hat{d}_R)	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	40	1:3	.005	.215	.008	.044	.008	.038	.010	.048	.008	.042	.009	.072	.010	.068
		1:1	.005	.024	.009	.005	.009	.007	.006	.006	.004	.005	.007	.013	.008	.015
		3:1	.	.036	.	.004	.	.011	.	.016	.	.025	.	.056	.	.072
	120	1:3	.003	.266	.005	.074	.005	.069	.002	.069	.002	.064	.004	.073	.004	.069
		1:1	.008	.017	.003	.004	.003	.004	.008	.002	.008	.001	.007	.011	.008	.012
		3:1	.	.035	.	.048	.	.057	.	.059	.	.068	.	.102	.	.113
Gleich	40	1:3	.002	.193	.006	.042	.004	.037	.015	.029	.015	.025	.008	.055	.010	.052
		1:1	.004	.016	.003	.005	.005	.005	.006	.002	.005	.003	.017	.020	.016	.021
		3:1	.	.041	.	.008	.	.012	.	.018	.	.026	.	.052	.	.064
	120	1:3	.002	.230	.001	.067	.002	.063	.003	.050	.003	.047	.004	.036	.005	.034
		1:1	.003	.010	.002	.005	.002	.005	.002	.005	.003	.005	.008	.014	.008	.014
		3:1	.	.033	.	.042	.	.048	.	.054	.	.061	.	.087	.	.097
g-h-1	40	1:3	.017	.250	.012	.035	.011	.027	.007	.066	.007	.058	.015	.099	.017	.090
		1:1	.018	.046	.017	.013	.016	.014	.006	.009	.005	.007	.013	.024	.015	.028
		3:1	.	.035	.	.005	.	.013	.	.019	.	.029	.	.069	.	.088
	120	1:3	.012	.327	.011	.082	.009	.075	.002	.097	.002	.092	.006	.126	.007	.119
		1:1	.019	.047	.010	.007	.009	.008	.004	.005	.004	.005	.008	.016	.008	.017
		3:1	.	.076	.	.068	.	.080	.	.075	.	.086	.	.127	.	.142
g-h-2	40	1:3	.021	.173	.005	.056	.007	.043	.010	.068	.009	.057	.040	.047	.046	.044
		1:1	.023	.021	.008	.016	.010	.018	.008	.005	.006	.004	.014	.019	.016	.023
		3:1	.	.040	.	.006	.	.007	.	.020	.	.031	.	.056	.	.069
	120	1:3	.020	.242	.005	.116	.005	.106	.006	.120	.005	.112	.023	.074	.023	.068
		1:1	.026	.023	.009	.012	.009	.012	.002	.004	.002	.003	.012	.014	.012	.015
		3:1	.	.036	.	.067	.	.079	.	.082	.	.094	.	.119	.	.132
g-h-3	40	1:3	.019	.228	.014	.032	.014	.023	.008	.063	.005	.055	.019	.086	.022	.078
		1:1	.022	.025	.018	.017	.017	.018	.005	.005	.005	.005	.014	.016	.017	.020
		3:1	.	.035	.	.005	.	.015	.	.018	.	.028	.	.067	.	.084
	120	1:3	.018	.316	.010	.090	.010	.083	.003	.112	.003	.104	.011	.123	.012	.116
		1:1	.024	.028	.010	.011	.010	.011	.006	.005	.006	.004	.012	.014	.012	.015
		3:1	.	.075	.	.079	.	.092	.	.082	.	.094	.	.129	.	.143
g-h-4	40	1:3	.014	.153	.002	.070	.003	.059	.008	.054	.007	.047	.031	.035	.036	.033
		1:1	.014	.025	.002	.009	.004	.011	.006	.007	.005	.006	.010	.018	.012	.020
		3:1	.	.042	.	.007	.	.004	.	.017	.	.027	.	.044	.	.057
	120	1:3	.011	.199	.002	.110	.002	.102	.003	.095	.003	.089	.017	.036	.019	.033
		1:1	.011	.029	.005	.010	.006	.012	.002	.005	.002	.004	.006	.016	.007	.016
		3:1	.	.041	.	.057	.	.067	.	.075	.	.085	.	.089	.	.100
Hyper	40	1:3	.060	.195	.009	.038	.010	.034	.024	.070	.023	.066	.044	.083	.047	.081
		1:1	.049	.014	.011	.007	.008	.006	.022	.008	.024	.009	.026	.032	.031	.037
		3:1	.	.038	.	.020	.	.029	.	.034	.	.044	.	.087	.	.103
	120	1:3	.077	.230	.005	.053	.006	.049	.023	.042	.022	.038	.030	.040	.031	.038
		1:1	.054	.013	.003	.003	.002	.004	.022	.013	.024	.014	.027	.045	.029	.048
		3:1	.	.067	.	.089	.	.098	.	.104	.	.112	.	.140	.	.150

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.1.4.5

Tabelle C-26. Mittlere Intervalllänge

			NC($d_{s,R}$)		Perc($d_{s,R}$)		Perc(\hat{d}_R)		BCA($d_{s,R}$)		BCA(\hat{d}_R)		Perc-t($d_{s,R}$)		Perc-t(\hat{d}_R)	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	40	1:3	1.690	1.813	1.946	3.480	1.879	3.360	1.961	3.459	1.894	3.339	1.818	2.987	1.755	2.884
		1:1	1.493	1.503	1.727	1.949	1.668	1.881	1.745	1.942	1.685	1.875	1.613	1.756	1.557	1.695
		3:1	.	1.650	.	1.398	.	1.350	.	1.399	.	1.351	.	1.317	.	1.271
	120	1:3	.950	.994	1.004	1.670	.993	1.652	1.005	1.665	.995	1.647	.987	1.600	.976	1.582
		1:1	.839	.839	.891	.958	.882	.948	.894	.956	.884	.946	.876	.933	.866	.923
		3:1	.	.933	.	.711	.	.703	.	.710	.	.702	.	.700	.	.693
Gleich	40	1:3	2.072	2.193	2.284	4.292	2.205	4.144	2.272	4.233	2.194	4.088	2.124	3.567	2.051	3.444
		1:1	1.821	1.826	2.097	2.359	2.025	2.278	2.078	2.308	2.007	2.228	1.933	2.090	1.866	2.018
		3:1	.	2.033	.	1.634	.	1.578	.	1.612	.	1.556	.	1.540	.	1.487
	120	1:3	1.166	1.206	1.195	2.008	1.182	1.986	1.192	1.988	1.179	1.967	1.173	1.906	1.160	1.885
		1:1	1.023	1.024	1.064	1.124	1.052	1.112	1.060	1.116	1.049	1.104	1.042	1.093	1.031	1.081
		3:1	.	1.151	.	.825	.	.816	.	.821	.	.812	.	.814	.	.806
g-h-1	40	1:3	1.230	1.366	1.531	2.711	1.478	2.618	1.576	2.716	1.521	2.622	1.457	2.457	1.407	2.372
		1:1	1.103	1.114	1.372	1.569	1.325	1.515	1.420	1.598	1.371	1.543	1.309	1.439	1.264	1.390
		3:1	.	1.187	.	1.152	.	1.113	.	1.172	.	1.131	.	1.080	.	1.043
	120	1:3	.690	.743	.778	1.317	.770	1.303	.784	1.320	.776	1.306	.769	1.285	.761	1.271
		1:1	.618	.619	.702	.781	.694	.772	.710	.784	.702	.776	.694	.763	.686	.754
		3:1	.	.669	.	.580	.	.574	.	.583	.	.576	.	.571	.	.564
g-h-2	40	1:3	1.454	1.507	1.814	2.650	1.752	2.558	1.797	2.646	1.735	2.555	1.640	2.226	1.583	2.149
		1:1	1.288	1.247	1.636	1.467	1.580	1.416	1.648	1.483	1.591	1.432	1.515	1.329	1.463	1.283
		3:1	.	1.371	.	1.093	.	1.056	.	1.098	.	1.060	.	1.039	.	1.004
	120	1:3	.812	.827	.920	1.210	.910	1.197	.919	1.212	.909	1.199	.897	1.163	.887	1.150
		1:1	.720	.698	.837	.704	.828	.697	.838	.704	.829	.697	.821	.689	.812	.681
		3:1	.	.776	.	.537	.	.531	.	.536	.	.530	.	.531	.	.525
g-h-3	40	1:3	1.171	1.276	1.477	2.456	1.426	2.371	1.518	2.462	1.466	2.377	1.398	2.212	1.350	2.136
		1:1	1.051	1.041	1.341	1.412	1.295	1.364	1.389	1.453	1.341	1.403	1.284	1.300	1.240	1.255
		3:1	.	1.110	.	1.044	.	1.008	.	1.068	.	1.032	.	.986	.	.952
	120	1:3	.655	.695	.748	1.167	.740	1.155	.753	1.170	.745	1.158	.738	1.137	.730	1.125
		1:1	.589	.580	.685	.691	.677	.684	.692	.696	.685	.688	.678	.676	.671	.669
		3:1	.	.626	.	.516	.	.511	.	.520	.	.514	.	.509	.	.504
g-h-4	40	1:3	1.830	1.878	2.238	3.286	2.160	3.172	2.171	3.211	2.096	3.101	1.996	2.711	1.928	2.617
		1:1	1.608	1.570	1.953	1.772	1.885	1.711	1.941	1.764	1.874	1.703	1.788	1.605	1.727	1.550
		3:1	.	1.753	.	1.296	.	1.251	.	1.290	.	1.246	.	1.235	.	1.192
	120	1:3	1.025	1.037	1.114	1.482	1.102	1.466	1.105	1.477	1.093	1.461	1.085	1.420	1.073	1.405
		1:1	.901	.882	.989	.837	.979	.828	.987	.835	.976	.826	.969	.820	.958	.812
		3:1	.	.994	.	.636	.	.629	.	.634	.	.627	.	.630	.	.623
Hyper	40	1:3	1.808	2.494	2.514	5.598	2.427	5.405	2.472	5.223	2.387	5.043	2.329	4.401	2.248	4.249
		1:1	1.566	2.044	2.064	3.165	1.993	3.056	2.056	3.041	1.985	2.936	1.924	2.544	1.858	2.457
		3:1	.	2.276	.	1.943	.	1.876	.	1.903	.	1.837	.	1.746	.	1.685
	120	1:3	1.009	1.351	1.366	2.593	1.352	2.565	1.358	2.535	1.343	2.508	1.317	2.371	1.303	2.345
		1:1	.879	1.144	1.096	1.253	1.084	1.240	1.082	1.245	1.070	1.232	1.073	1.216	1.061	1.203
		3:1	.	1.290	.	.866	.	.856	.	.865	.	.856	.	.856	.	.846

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.1.4.5

Tabelle C-27. $|CP - .95|$

			NC($d'_{s,R}$)		Perc($d'_{s,R}$)		Perc(\hat{d}'_R)		BCA($d'_{s,R}$)		BCA(\hat{d}'_R)		Perc-t($d'_{s,R}$)		Perc-t(\hat{d}'_R)	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	40	1:3	.006	.009	.007	.009	.009	.009	.014	.017	.014	.017	.023	.012	.024	.020
		1:1	.003	.009	.009	.005	.009	.008	.006	.008	.005	.006	.006	.008	.007	.013
		3:1	.	.005	.	.006	.	.003	.	.004	.	.011	.	.010	.	.034
	120	1:3	.006	.003	.007	.003	.007	.004	.003	.004	.002	.004	.009	.013	.009	.013
		1:1	.008	.005	.003	.004	.003	.005	.009	.002	.009	.001	.008	.008	.009	.010
		3:1	.	.004	.	.007	.	.006	.	.002	.	.002	.	.006	.	.010
Gleich	40	1:3	.011	.013	.005	.009	.004	.007	.013	.007	.015	.009	.014	.008	.017	.008
		1:1	.003	.005	.003	.005	.005	.006	.006	.004	.006	.004	.011	.017	.012	.020
		3:1	.	.004	.	.004	.	.004	.	.006	.	.013	.	.011	.	.028
	120	1:3	.006	.005	.002	.004	.003	.002	.003	.004	.003	.004	.006	.012	.006	.012
		1:1	.004	.005	.002	.005	.002	.006	.002	.005	.002	.005	.007	.012	.007	.012
		3:1	.	.002	.	.002	.	.001	.	.004	.	.005	.	.007	.	.012
g-h-1	40	1:3	.006	.018	.013	.014	.012	.016	.015	.021	.016	.024	.030	.030	.031	.040
		1:1	.015	.026	.017	.013	.016	.014	.007	.011	.007	.008	.019	.021	.020	.029
		3:1	.	.017	.	.016	.	.005	.	.003	.	.010	.	.021	.	.048
	120	1:3	.002	.021	.012	.009	.011	.010	.004	.008	.004	.008	.015	.031	.013	.032
		1:1	.017	.024	.010	.007	.009	.008	.004	.005	.004	.005	.009	.013	.010	.015
		3:1	.	.022	.	.010	.	.007	.	.004	.	.005	.	.012	.	.019
g-h-2	40	1:3	.005	.026	.012	.013	.014	.017	.008	.021	.011	.021	.018	.023	.028	.026
		1:1	.019	.026	.008	.016	.011	.019	.008	.007	.007	.006	.022	.019	.025	.026
		3:1	.	.022	.	.016	.	.011	.	.002	.	.011	.	.016	.	.042
	120	1:3	.012	.031	.011	.014	.011	.016	.005	.009	.005	.009	.008	.021	.010	.023
		1:1	.024	.029	.009	.012	.009	.013	.002	.005	.002	.004	.014	.013	.015	.016
		3:1	.	.023	.	.012	.	.011	.	.005	.	.004	.	.010	.	.016
g-h-3	40	1:3	.003	.005	.018	.019	.015	.021	.013	.020	.014	.024	.025	.023	.026	.033
		1:1	.018	.011	.018	.017	.017	.017	.005	.007	.005	.006	.021	.013	.023	.021
		3:1	.	.008	.	.020	.	.008	.	.003	.	.009	.	.014	.	.042
	120	1:3	.007	.006	.013	.014	.013	.014	.006	.007	.004	.007	.013	.022	.013	.023
		1:1	.022	.012	.010	.011	.010	.011	.006	.005	.006	.004	.013	.011	.013	.013
		3:1	.	.010	.	.011	.	.006	.	.003	.	.003	.	.009	.	.017
g-h-4	40	1:3	.007	.032	.003	.010	.007	.014	.009	.017	.011	.017	.014	.020	.025	.023
		1:1	.011	.027	.002	.009	.005	.013	.007	.008	.006	.007	.011	.016	.013	.019
		3:1	.	.025	.	.008	.	.010	.	.004	.	.009	.	.014	.	.026
	120	1:3	.016	.031	.005	.008	.007	.009	.004	.008	.005	.007	.006	.013	.006	.013
		1:1	.010	.032	.005	.010	.006	.012	.002	.005	.002	.004	.010	.014	.010	.015
		3:1	.	.023	.	.005	.	.006	.	.006	.	.007	.	.007	.	.010
Hyper	40	1:3	.022	.025	.011	.012	.009	.010	.024	.055	.027	.055	.037	.053	.041	.056
		1:1	.040	.005	.011	.007	.007	.006	.023	.009	.026	.011	.024	.030	.029	.037
		3:1	.	.007	.	.004	.	.013	.	.008	.	.025	.	.037	.	.065
	120	1:3	.022	.003	.009	.005	.009	.004	.020	.010	.020	.012	.025	.040	.028	.041
		1:1	.050	.011	.003	.003	.002	.005	.022	.012	.024	.015	.027	.041	.029	.047
		3:1	.	.011	.	.005	.	.010	.	.015	.	.024	.	.031	.	.045

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.1.4.5

Tabelle C-28. Mittlere Intervalllänge

			NC($d'_{s,R}$)		Perc($d'_{s,R}$)		Perc(\hat{d}'_R)		BCA($d'_{s,R}$)		BCA(\hat{d}'_R)		Perc-t($d'_{s,R}$)		Perc-t(\hat{d}'_R)	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	40	1:3	1.813	2.317	2.051	3.299	1.876	3.092	2.072	3.286	1.907	3.077	1.900	2.759	1.768	2.628
		1:1	1.505	1.575	1.727	1.949	1.657	1.858	1.746	1.957	1.680	1.863	1.588	1.771	1.530	1.683
		3:1	.	1.378	.	1.612	.	1.428	.	1.624	.	1.435	.	1.466	.	1.289
	120	1:3	1.016	1.245	1.036	1.413	1.015	1.397	1.039	1.412	1.018	1.396	1.014	1.318	.995	1.303
		1:1	.841	.877	.891	.958	.881	.943	.894	.958	.884	.943	.873	.934	.863	.919
		3:1	.	.766	.	.827	.	.799	.	.828	.	.800	.	.808	.	.780
Gleich	40	1:3	2.168	2.765	2.459	4.119	2.237	3.854	2.441	4.046	2.234	3.793	2.245	3.258	2.093	3.106
		1:1	1.830	1.893	2.097	2.359	2.012	2.247	2.080	2.323	1.999	2.210	1.901	2.092	1.831	1.983
		3:1	.	1.653	.	1.909	.	1.681	.	1.887	.	1.661	.	1.716	.	1.503
	120	1:3	1.221	1.489	1.238	1.676	1.214	1.657	1.233	1.659	1.209	1.640	1.203	1.538	1.181	1.521
		1:1	1.024	1.057	1.064	1.124	1.052	1.106	1.060	1.117	1.048	1.099	1.040	1.089	1.028	1.071
		3:1	.	.926	.	.969	.	.935	.	.965	.	.931	.	.946	.	.912
g-h-1	40	1:3	1.384	1.802	1.617	2.623	1.476	2.450	1.660	2.665	1.523	2.466	1.573	2.412	1.458	2.295
		1:1	1.120	1.196	1.372	1.569	1.318	1.499	1.422	1.615	1.370	1.539	1.285	1.477	1.241	1.407
		3:1	.	1.048	.	1.316	.	1.170	.	1.357	.	1.200	.	1.215	.	1.069
	120	1:3	.774	.960	.817	1.151	.799	1.138	.824	1.163	.806	1.149	.816	1.106	.799	1.094
		1:1	.621	.665	.702	.781	.694	.770	.710	.786	.702	.775	.690	.769	.683	.758
		3:1	.	.575	.	.668	.	.647	.	.673	.	.651	.	.654	.	.632
g-h-2	40	1:3	1.592	1.910	1.826	2.495	1.685	2.335	1.847	2.515	1.713	2.350	1.671	2.036	1.590	1.945
		1:1	1.308	1.306	1.636	1.467	1.567	1.395	1.650	1.495	1.586	1.421	1.483	1.331	1.430	1.261
		3:1	.	1.152	.	1.277	.	1.118	.	1.288	.	1.129	.	1.160	.	1.003
	120	1:3	.886	1.032	.878	1.020	.863	1.008	.882	1.025	.869	1.013	.856	.956	.845	.945
		1:1	.724	.729	.837	.704	.827	.693	.838	.706	.828	.694	.815	.688	.806	.677
		3:1	.	.640	.	.637	.	.613	.	.638	.	.614	.	.624	.	.599
g-h-3	40	1:3	1.332	1.679	1.551	2.370	1.419	2.215	1.599	2.416	1.471	2.239	1.513	2.157	1.412	2.053
		1:1	1.070	1.119	1.341	1.412	1.288	1.348	1.390	1.469	1.339	1.399	1.259	1.331	1.217	1.267
		3:1	.	.980	.	1.201	.	1.064	.	1.238	.	1.095	.	1.114	.	.974
	120	1:3	.743	.897	.769	1.017	.753	1.006	.776	1.031	.760	1.019	.767	.976	.752	.965
		1:1	.592	.623	.685	.691	.677	.681	.693	.698	.685	.688	.674	.681	.667	.671
		3:1	.	.538	.	.600	.	.580	.	.606	.	.586	.	.589	.	.568
g-h-4	40	1:3	1.945	2.330	2.222	3.057	2.055	2.858	2.185	3.020	2.033	2.824	1.968	2.409	1.874	2.305
		1:1	1.627	1.624	1.953	1.772	1.866	1.682	1.944	1.775	1.863	1.685	1.742	1.592	1.676	1.505
		3:1	.	1.434	.	1.512	.	1.323	.	1.510	.	1.322	.	1.367	.	1.184
	120	1:3	1.085	1.273	1.046	1.221	1.030	1.206	1.043	1.220	1.028	1.205	1.012	1.138	1.000	1.125
		1:1	.904	.909	.989	.837	.977	.823	.987	.836	.975	.821	.961	.817	.951	.803
		3:1	.	.801	.	.757	.	.728	.	.756	.	.727	.	.742	.	.713
Hyper	40	1:3	2.023	3.186	2.494	5.215	2.331	4.918	2.461	4.840	2.312	4.571	2.363	3.856	2.255	3.679
		1:1	1.603	2.124	2.064	3.165	1.978	3.007	2.055	3.074	1.973	2.912	1.927	2.533	1.849	2.385
		3:1	.	1.779	.	2.245	.	1.977	.	2.221	.	1.942	.	1.929	.	1.663
	120	1:3	1.124	1.682	1.277	2.096	1.256	2.069	1.274	2.050	1.255	2.024	1.230	1.814	1.213	1.791
		1:1	.895	1.182	1.096	1.253	1.081	1.230	1.083	1.248	1.067	1.226	1.073	1.205	1.059	1.183
		3:1	.	.999	.	1.039	.	.998	.	1.037	.	.997	.	1.015	.	.975

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.1.4.5

Tabelle C-29. $|CP - .95|$

			NC($d_{s,R}^{S_1}$)		Perc($d_{s,R}^{S_1}$)		Perc($\hat{d}_R^{S_1}$)		BCA($d_{s,R}^{S_1}$)		BCA($\hat{d}_R^{S_1}$)		Perc-t($d_{s,R}^{S_1}$)		Perc-t($\hat{d}_R^{S_1}$)	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	40	1:3	.008	.004	.008	.009	.016	.012	.005	.008	.005	.007	.018	.024	.039	.025
		1:1	.006	.005	.010	.005	.013	.009	.003	.004	.004	.002	.016	.013	.022	.018
		3:1	.	.006	.	.006	.	.007	.	.003	.	.002	.	.013	.	.017
	120	1:3	.006	.007	.010	.003	.012	.006	.003	.003	.004	.003	.011	.012	.018	.016
		1:1	.009	.005	.005	.004	.006	.005	.006	.002	.006	.003	.010	.008	.013	.011
		3:1	.	.004	.	.007	.	.007	.	.002	.	.001	.	.007	.	.007
Gleich	40	1:3	.012	.007	.004	.009	.012	.011	.003	.003	.005	.005	.025	.036	.042	.032
		1:1	.007	.004	.003	.004	.008	.007	.002	.002	.001	.003	.015	.023	.020	.027
		3:1	.	.002	.	.005	.	.005	.	.005	.	.004	.	.012	.	.016
	120	1:3	.005	.003	.003	.003	.005	.004	.004	.005	.004	.005	.016	.015	.022	.020
		1:1	.004	.005	.002	.005	.004	.005	.002	.005	.002	.005	.009	.014	.010	.014
		3:1	.	.003	.	.002	.	.001	.	.004	.	.003	.	.008	.	.008
g-h-1	40	1:3	.022	.019	.016	.014	.025	.021	.007	.006	.009	.008	.022	.016	.051	.035
		1:1	.020	.022	.018	.012	.018	.015	.004	.004	.004	.006	.028	.011	.038	.024
		3:1	.	.019	.	.013	.	.014	.	.003	.	.003	.	.025	.	.030
	120	1:3	.029	.030	.013	.010	.015	.011	.003	.006	.003	.004	.015	.020	.022	.024
		1:1	.022	.024	.011	.007	.012	.009	.002	.003	.003	.005	.015	.011	.017	.014
		3:1	.	.021	.	.009	.	.010	.	.002	.	.002	.	.012	.	.014
g-h-2	40	1:3	.027	.025	.013	.011	.027	.023	.005	.013	.005	.009	.028	.021	.060	.047
		1:1	.019	.030	.015	.016	.020	.020	.004	.004	.006	.003	.020	.021	.030	.034
		3:1	.	.028	.	.017	.	.019	.	.003	.	.002	.	.022	.	.029
	120	1:3	.020	.031	.014	.014	.020	.019	.005	.008	.005	.007	.021	.021	.031	.027
		1:1	.017	.031	.014	.012	.015	.014	.002	.004	.001	.004	.018	.017	.022	.020
		3:1	.	.030	.	.012	.	.012	.	.003	.	.004	.	.011	.	.014
g-h-3	40	1:3	.018	.006	.018	.019	.026	.025	.003	.007	.006	.008	.029	.016	.055	.030
		1:1	.016	.013	.022	.017	.023	.018	.007	.003	.008	.003	.025	.013	.035	.023
		3:1	.	.009	.	.019	.	.019	.	.006	.	.007	.	.021	.	.025
	120	1:3	.015	.010	.017	.015	.019	.016	.003	.006	.004	.004	.014	.016	.022	.023
		1:1	.014	.014	.013	.011	.012	.011	.003	.003	.004	.003	.015	.010	.018	.013
		3:1	.	.011	.	.010	.	.010	.	.004	.	.004	.	.013	.	.014
g-h-4	40	1:3	.033	.031	.004	.011	.023	.021	.004	.010	.006	.008	.028	.023	.044	.038
		1:1	.027	.029	.010	.009	.017	.014	.004	.007	.005	.005	.017	.020	.023	.022
		3:1	.	.030	.	.010	.	.012	.	.003	.	.002	.	.017	.	.019
	120	1:3	.028	.033	.011	.008	.018	.012	.003	.008	.002	.007	.019	.017	.026	.024
		1:1	.025	.034	.012	.010	.014	.014	.003	.005	.003	.004	.010	.016	.012	.017
		3:1	.	.028	.	.006	.	.007	.	.005	.	.005	.	.008	.	.010
Hyper	40	1:3	.011	.021	.012	.012	.018	.011	.038	.051	.038	.055	.039	.052	.078	.072
		1:1	.006	.006	.016	.007	.015	.006	.018	.008	.020	.011	.022	.034	.038	.043
		3:1	.	.009	.	.005	.	.003	.	.008	.	.010	.	.039	.	.043
	120	1:3	.012	.004	.013	.005	.013	.003	.027	.011	.029	.015	.031	.042	.045	.055
		1:1	.023	.012	.010	.004	.007	.006	.033	.013	.037	.016	.042	.045	.048	.051
		3:1	.	.013	.	.006	.	.008	.	.016	.	.020	.	.036	.	.041

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.1.4.5

Tabelle C-30. Mittlere Intervalllänge

			NC($d_{s,R}^{S_1}$)		Perc($d_{s,R}^{S_1}$)		Perc($\hat{d}_R^{S_1}$)		BCA($d_{s,R}^{S_1}$)		BCA($\hat{d}_R^{S_1}$)		Perc-t($d_{s,R}^{S_1}$)		Perc-t($\hat{d}_R^{S_1}$)	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	40	1:3	2.319	2.498	5.973	6.318	5.030	5.321	5.600	6.091	4.853	5.262	2.963	3.402	2.495	2.865
		1:1	1.687	1.652	2.285	2.261	2.125	2.104	2.324	2.282	2.164	2.129	1.851	1.904	1.722	1.771
		3:1	.	1.412	.	1.723	.	1.645	.	1.722	.	1.645	.	1.505	.	1.438
	120	1:3	1.141	1.273	1.418	1.562	1.355	1.492	1.423	1.563	1.361	1.496	1.273	1.401	1.216	1.338
		1:1	.909	.903	1.016	1.010	.994	.988	1.016	1.008	.994	.987	.963	.971	.942	.950
		3:1	.	.787	.	.855	.	.843	.	.854	.	.842	.	.827	.	.816
Gleich	40	1:3	2.721	2.959	7.179	8.030	6.046	6.762	7.277	7.791	6.279	7.126	3.303	3.881	2.781	3.269
		1:1	2.000	1.959	2.642	2.665	2.457	2.479	2.590	2.609	2.411	2.432	2.118	2.197	1.971	2.044
		3:1	.	1.679	.	1.985	.	1.896	.	1.957	.	1.869	.	1.741	.	1.663
	120	1:3	1.341	1.515	1.604	1.819	1.532	1.737	1.576	1.794	1.507	1.716	1.409	1.603	1.346	1.531
		1:1	1.084	1.079	1.158	1.164	1.133	1.139	1.149	1.156	1.124	1.131	1.103	1.115	1.079	1.091
		3:1	.	.943	.	.985	.	.971	.	.981	.	.967	.	.958	.	.944
g-h-1	40	1:3	1.864	1.984	5.283	5.495	4.449	4.628	2.782	3.142	2.342	2.646
		1:1	1.313	1.281	1.963	1.897	1.826	1.764	2.062	1.982	1.922	1.851	1.607	1.634	1.494	1.520
		3:1	.	1.085	.	1.459	.	1.394	.	1.474	.	1.408	.	1.253	.	1.197
	120	1:3	.910	.990	1.248	1.318	1.192	1.259	1.283	1.347	1.228	1.289	1.157	1.213	1.105	1.159
		1:1	.702	.696	.856	.846	.838	.827	.863	.851	.844	.833	.811	.820	.793	.803
		3:1	.	.600	.	.706	.	.696	.	.708	.	.698	.	.679	.	.669
g-h-2	40	1:3	2.125	2.096	5.644	5.145	4.753	4.333	.	5.448	.	4.749	2.583	2.550	2.175	2.148
		1:1	1.515	1.379	2.042	1.732	1.900	1.611	2.090	1.767	1.947	1.648	1.596	1.412	1.485	1.313
		3:1	.	1.186	.	1.352	.	1.292	.	1.352	.	1.291	.	1.147	.	1.096
	120	1:3	1.023	1.060	1.201	1.158	1.147	1.106	1.197	1.160	1.145	1.110	1.057	1.028	1.010	.982
		1:1	.802	.752	.854	.742	.835	.726	.851	.740	.832	.725	.799	.710	.782	.695
		3:1	.	.657	.	.641	.	.632	.	.639	.	.630	.	.616	.	.607
g-h-3	40	1:3	1.808	1.844	5.111	4.854	4.304	4.087	2.641	2.795	2.224	2.353
		1:1	1.263	1.199	1.856	1.699	1.727	1.580	1.978	1.801	1.845	1.685	1.508	1.462	1.402	1.360
		3:1	.	1.017	.	1.312	.	1.253	.	1.337	.	1.277	.	1.124	.	1.074
	120	1:3	.879	.924	1.170	1.167	1.117	1.115	1.205	1.197	1.154	1.146	1.081	1.071	1.032	1.023
		1:1	.675	.652	.800	.747	.783	.731	.809	.754	.791	.738	.758	.725	.741	.709
		3:1	.	.562	.	.626	.	.617	.	.629	.	.620	.	.603	.	.594
g-h-4	40	1:3	2.543	2.556	7.292	7.057	6.141	5.943	6.688	10.123	5.722	7.537	2.973	2.960	2.504	2.493
		1:1	1.837	1.698	2.396	2.117	2.229	1.969	2.381	2.109	2.216	1.965	1.818	1.669	1.691	1.552
		3:1	.	1.466	.	1.596	.	1.525	.	1.583	.	1.513	.	1.364	.	1.303
	120	1:3	1.220	1.302	1.331	1.366	1.271	1.305	1.308	1.355	1.250	1.296	1.137	1.193	1.086	1.139
		1:1	.976	.929	.965	.865	.944	.846	.959	.861	.938	.842	.903	.826	.883	.808
		3:1	.	.816	.	.756	.	.745	.	.753	.	.742	.	.730	.	.719
Hyper	40	1:3	2.525	3.439	3.842	7.744	3.235	6.521	3.528	7.023	2.998	5.997	3.055	4.722	2.572	3.976
		1:1	1.804	2.170	2.473	3.955	2.301	3.679	2.488	3.786	2.317	3.532	2.013	2.669	1.872	2.482
		3:1	.	1.802	.	2.470	.	2.360	.	2.407	.	2.300	.	1.953	.	1.865
	120	1:3	1.244	1.694	1.600	2.364	1.529	2.258	1.607	2.283	1.537	2.185	1.407	1.860	1.344	1.777
		1:1	.971	1.192	1.126	1.274	1.102	1.246	1.104	1.269	1.080	1.242	1.060	1.216	1.038	1.189
		3:1	.	1.011	.	1.043	.	1.028	.	1.042	.	1.027	.	1.015	.	1.001

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.1.4.5

Tabelle C-31. $|CP - .95|$

			NC($d_{s,R}^{S_2}$)		Perc($d_{s,R}^{S_2}$)		Perc($\hat{d}_R^{S_2}$)		BCA($d_{s,R}^{S_2}$)		BCA($\hat{d}_R^{S_2}$)		Perc-t($d_{s,R}^{S_2}$)		Perc-t($\hat{d}_R^{S_2}$)	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	40	1:3	.016	.038	.008	.008	.010	.009	.006	.012	.006	.012	.030	.053	.031	.051
		1:1	.006	.007	.010	.012	.012	.014	.003	.007	.004	.007	.013	.010	.020	.015
		3:1	.	.013	.	.016	.	.026	.	.007	.	.011	.	.019	.	.047
	120	1:3	.007	.011	.005	.003	.005	.003	.002	.004	.001	.004	.010	.015	.010	.014
		1:1	.010	.007	.004	.006	.006	.008	.007	.001	.006	.001	.012	.007	.014	.008
		3:1	.	.010	.	.013	.	.014	.	.003	.	.002	.	.014	.	.021
Gleich	40	1:3	.020	.042	.004	.009	.004	.009	.013	.015	.013	.014	.026	.048	.027	.047
		1:1	.006	.006	.004	.005	.007	.009	.003	.004	.002	.003	.013	.003	.018	.003
		3:1	.	.019	.	.006	.	.019	.	.002	.	.009	.	.018	.	.037
	120	1:3	.004	.014	.001	.002	.002	.002	.003	.003	.003	.003	.006	.015	.006	.015
		1:1	.004	.005	.003	.005	.004	.005	.002	.003	.002	.003	.007	.003	.008	.003
		3:1	.	.007	.	.005	.	.009	.	.003	.	.002	.	.012	.	.018
g-h-1	40	1:3	.026	.037	.013	.014	.013	.014	.007	.010	.008	.009	.037	.042	.039	.040
		1:1	.023	.027	.017	.018	.018	.019	.004	.010	.006	.010	.029	.023	.038	.030
		3:1	.	.027	.	.023	.	.031	.	.011	.	.013	.	.022	.	.053
	120	1:3	.018	.018	.008	.007	.008	.006	.002	.003	.003	.003	.013	.016	.014	.015
		1:1	.028	.024	.009	.014	.009	.012	.003	.005	.003	.005	.020	.010	.022	.012
		3:1	.	.033	.	.019	.	.020	.	.005	.	.005	.	.022	.	.029
g-h-2	40	1:3	.069	.067	.002	.005	.005	.007	.004	.008	.003	.008	.068	.059	.071	.057
		1:1	.073	.045	.003	.005	.003	.008	.003	.007	.003	.006	.062	.028	.070	.032
		3:1	.	.059	.	.009	.	.014	.	.011	.	.008	.	.041	.	.059
	120	1:3	.061	.039	.003	.004	.004	.004	.004	.003	.004	.003	.031	.018	.033	.019
		1:1	.085	.049	.005	.005	.006	.005	.002	.003	.003	.004	.036	.015	.038	.016
		3:1	.	.072	.	.003	.	.005	.	.004	.	.003	.	.038	.	.040
g-h-3	40	1:3	.040	.048	.012	.014	.013	.015	.008	.012	.008	.012	.043	.046	.044	.046
		1:1	.044	.036	.013	.017	.015	.019	.003	.007	.004	.009	.045	.024	.051	.027
		3:1	.	.038	.	.020	.	.029	.	.012	.	.014	.	.027	.	.056
	120	1:3	.036	.027	.007	.007	.007	.007	.003	.003	.003	.004	.019	.017	.019	.017
		1:1	.047	.034	.006	.012	.008	.013	.004	.006	.004	.006	.026	.013	.028	.014
		3:1	.	.049	.	.014	.	.015	.	.004	.	.004	.	.027	.	.030
g-h-4	40	1:3	.057	.073	.002	.008	.003	.009	.005	.014	.005	.014	.061	.064	.066	.065
		1:1	.056	.039	.003	.004	.002	.006	.003	.005	.003	.005	.047	.021	.056	.027
		3:1	.	.048	.	.004	.	.011	.	.008	.	.010	.	.038	.	.061
	120	1:3	.045	.033	.002	.003	.002	.003	.003	.005	.003	.005	.025	.019	.026	.020
		1:1	.064	.035	.003	.005	.003	.006	.002	.003	.003	.003	.022	.010	.027	.010
		3:1	.	.060	.	.003	.	.003	.	.001	.	.002	.	.030	.	.036
Hyper	40	1:3	.106	.060	.007	.012	.006	.012	.025	.078	.025	.078	.104	.109	.107	.107
		1:1	.105	.019	.005	.008	.009	.009	.037	.049	.037	.042	.119	.054	.130	.050
		3:1	.	.019	.	.008	.	.014	.	.052	.	.043	.	.109	.	.097
	120	1:3	.116	.016	.007	.007	.006	.007	.017	.025	.018	.025	.071	.019	.073	.018
		1:1	.135	.023	.005	.011	.004	.011	.019	.032	.019	.032	.087	.035	.093	.036
		3:1	.	.041	.	.011	.	.012	.	.054	.	.046	.	.089	.	.090

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.1.4.5

Tabelle C-32. Mittlere Intervalllänge

			NC($d_{s,R}^{S_2}$)		Perc($d_{s,R}^{S_2}$)		Perc($\hat{d}_R^{S_2}$)		BCA($d_{s,R}^{S_2}$)		BCA($\hat{d}_R^{S_2}$)		Perc-t($d_{s,R}^{S_2}$)		Perc-t($\hat{d}_R^{S_2}$)	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	40	1:3	1.741	2.006	2.109	2.422	2.015	2.314	2.117	2.440	2.022	2.329	1.949	2.491	1.862	2.379
		1:1	1.688	1.710	2.293	2.313	2.133	2.152	2.324	2.366	2.164	2.202	1.852	2.062	1.723	1.918
		3:1	.	2.081	.	5.461	.	4.599	2.934	.	2.471
	120	1:3	.971	1.123	1.045	1.202	1.030	1.185	1.045	1.203	1.030	1.186	1.016	1.194	1.002	1.177
		1:1	.912	.911	1.018	1.017	.996	.996	1.018	1.017	.996	.995	.966	.987	.945	.966
		3:1	.	.987	.	1.254	.	1.198	.	1.261	.	1.207	.	1.131	.	1.081
Gleich	40	1:3	2.066	2.371	2.335	2.618	2.230	2.501	2.318	2.608	2.213	2.490	2.248	2.909	2.147	2.779
		1:1	2.002	2.021	2.640	2.578	2.456	2.398	2.594	2.534	2.415	2.356	2.116	2.277	1.969	2.118
		3:1	.	2.393	.	6.383	.	5.375	.	5.540	.	4.715	.	2.961	.	2.493
	120	1:3	1.174	1.362	1.206	1.387	1.189	1.367	1.202	1.384	1.185	1.365	1.192	1.421	1.175	1.401
		1:1	1.087	1.089	1.161	1.149	1.136	1.125	1.152	1.140	1.127	1.116	1.106	1.120	1.082	1.096
		3:1	.	1.138	.	1.355	.	1.294	.	1.323	.	1.265	.	1.189	.	1.136
g-h-1	40	1:3	1.313	1.506	1.804	2.101	1.724	2.007	1.834	2.137	1.752	2.040	1.616	2.031	1.544	1.940
		1:1	1.314	1.333	1.967	2.015	1.830	1.874	2.067	2.135	1.926	1.987	1.606	1.921	1.494	1.787
		3:1	.	1.725	.	5.083	.	4.280	3.092	.	2.604
	120	1:3	.721	.822	.838	.948	.826	.935	.841	.952	.829	.939	.809	.924	.798	.911
		1:1	.702	.704	.857	.865	.839	.847	.864	.873	.846	.854	.811	.853	.794	.834
		3:1	.	.813	.	1.162	.	1.110	.	1.195	.	1.143	.	1.093	.	1.044
g-h-2	40	1:3	1.551	1.723	2.399	2.441	2.291	2.331	2.283	2.343	2.180	2.237	2.156	2.618	2.059	2.501
		1:1	1.550	1.481	2.829	2.428	2.632	2.259	2.683	2.323	2.500	2.162	2.164	2.254	2.013	2.097
		3:1	.	1.889	.	6.278	.	5.286	.	10.291	.	8.865	.	3.701	.	3.117
	120	1:3	.845	.945	1.118	1.126	1.103	1.110	1.101	1.113	1.085	1.097	1.063	1.129	1.048	1.113
		1:1	.807	.771	1.192	1.017	1.166	.995	1.171	1.002	1.146	.980	1.088	.995	1.064	.973
		3:1	.	.840	.	1.336	.	1.276	.	1.301	.	1.244	.	1.217	.	1.162
g-h-3	40	1:3	1.264	1.409	1.834	2.028	1.752	1.937	1.839	2.045	1.756	1.952	1.673	2.003	1.599	1.913
		1:1	1.273	1.255	2.035	1.963	1.893	1.826	2.106	2.049	1.963	1.908	1.683	1.922	1.565	1.788
		3:1	.	1.633	.	4.898	.	4.125	.	3.232	.	2.767	.	3.126	.	2.632
	120	1:3	.688	.772	.847	.913	.835	.900	.847	.915	.835	.902	.821	.895	.809	.882
		1:1	.678	.661	.890	.841	.871	.823	.894	.846	.874	.828	.842	.838	.824	.820
		3:1	.	.762	.	1.129	.	1.078	.	1.152	.	1.103	.	1.072	.	1.024
g-h-4	40	1:3	1.906	2.139	2.859	2.845	2.731	2.717	2.709	2.732	2.587	2.608	2.484	3.208	2.373	3.064
		1:1	1.883	1.819	3.597	3.029	3.346	2.818	3.367	2.845	3.138	2.647	2.515	2.592	2.339	2.411
		3:1	.	2.288	.	9.177	.	7.728	.	12.031	.	10.492	.	4.600	.	3.874
	120	1:3	1.046	1.194	1.316	1.354	1.298	1.335	1.294	1.338	1.276	1.319	1.244	1.389	1.226	1.370
		1:1	.980	.953	1.396	1.204	1.365	1.178	1.367	1.183	1.338	1.157	1.251	1.157	1.224	1.132
		3:1	.	1.001	.	1.585	.	1.514	.	1.527	.	1.461	.	1.340	.	1.280
Hyper	40	1:3	1.897	2.915	2.669	2.963	2.549	2.830	2.595	2.806	2.478	2.680	3.214	3.739	3.071	3.572
		1:1	1.848	2.436	2.810	2.710	2.614	2.521	2.655	2.621	2.470	2.437	3.027	3.034	2.815	2.822
		3:1	.	2.575	.	2.667	.	2.246	.	2.554	.	2.148	.	3.064	.	2.581
	120	1:3	1.050	1.638	1.564	1.824	1.542	1.798	1.560	1.768	1.538	1.743	1.598	1.838	1.575	1.812
		1:1	.996	1.284	1.587	1.570	1.553	1.536	1.566	1.497	1.532	1.464	1.680	1.477	1.644	1.445
		3:1	.	1.306	.	1.635	.	1.562	.	1.536	.	1.467	.	1.572	.	1.502

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.1.4.5

Anhang D Ergebnisse der Studie 2

Tabelle D-1. \overline{B}_{MC}^2

			$d_{D,s}$		\hat{d}_D		$d_{D,s}^{Boot}$		\hat{d}_D^{Boot}	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	.002	.002	.000	.000	.001	.001	.004	.004
		.4	.002	.002	.000	.000	.001	.001	.004	.004
		.8	.002	.002	.000	.000	.001	.001	.004	.004
	60	.0	.000	.000	.000	.000	.000	.000	.000	.000
		.4	.000	.000	.000	.000	.000	.000	.000	.000
		.8	.000	.000	.000	.000	.000	.000	.000	.000
Gleich	20	.0	.001	.001	.000	.000	.001	.001	.004	.004
		.4	.001	.001	.000	.000	.001	.001	.005	.004
		.8	.003	.000	.000	.000	.001	.001	.004	.005
	60	.0	.000	.000	.000	.000	.000	.000	.000	.000
		.4	.000	.000	.000	.000	.000	.000	.000	.000
		.8	.000	.000	.000	.000	.000	.000	.000	.000
g-h-1	20	.0	.019	.027	.009	.014	.001	.002	.000	.000
		.4	.025	.034	.013	.019	.003	.004	.000	.000
		.8	.033	.044	.018	.027	.004	.006	.001	.001
	60	.0	.004	.005	.002	.004	.000	.001	.000	.000
		.4	.005	.007	.003	.005	.001	.001	.000	.000
		.8	.009	.010	.006	.007	.002	.002	.001	.001
g-h-2	20	.0	.008	.005	.002	.002	.000	.000	.003	.002
		.4	.010	.008	.003	.003	.000	.000	.003	.002
		.8	.014	.012	.006	.005	.000	.000	.002	.001
	60	.0	.001	.001	.000	.000	.000	.000	.000	.000
		.4	.002	.001	.001	.000	.000	.000	.000	.000
		.8	.001	.001	.001	.001	.000	.000	.000	.000
g-h-3	20	.0	.033	.033	.019	.018	.004	.004	.000	.001
		.4	.048	.049	.029	.030	.009	.009	.003	.003
		.8	.076	.061	.051	.039	.019	.012	.009	.005
	60	.0	.008	.008	.006	.006	.002	.002	.001	.001
		.4	.013	.014	.010	.011	.003	.004	.002	.002
		.8	.024	.018	.020	.014	.008	.005	.006	.003
g-h-4	20	.0	.002	.001	.000	.001	.001	.001	.004	.004
		.4	.003	.001	.000	.001	.001	.001	.005	.004
		.8	.006	.001	.001	.001	.001	.001	.004	.005
	60	.0	.000	.000	.000	.000	.000	.000	.000	.000
		.4	.000	.000	.000	.000	.000	.000	.000	.000
		.8	.001	.000	.000	.000	.000	.000	.000	.000
Hyper	20	.0	.002	.000	.000	.000	.000	.001	.003	.004
		.4	.002	.000	.000	.000	.000	.001	.004	.004
		.8	.001	.000	.000	.001	.000	.001	.002	.005
	60	.0	.000	.000	.000	.000	.000	.000	.000	.000
		.4	.000	.000	.000	.000	.000	.000	.000	.001
		.8	.000	.000	.000	.000	.000	.000	.000	.000

Anmerkung: Vorpalte 1 gibt die Verteilung an, Vorpalte 2 den Stichprobenumfang N , Vorpalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-2. MSE_{MC}

			$d_{D,s}$		\hat{d}_D		$d_{D,s}^{Boot}$		\hat{d}_D^{Boot}	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	.149	.147	.135	.134	.129	.128	.122	.121
		.4	.098	.119	.089	.108	.085	.103	.082	.099
		.8	.056	.090	.050	.081	.048	.078	.048	.076
	60	.0	.044	.044	.043	.042	.042	.042	.041	.041
		.4	.030	.036	.029	.035	.028	.034	.028	.034
		.8	.016	.027	.015	.026	.015	.026	.015	.025
Gleich	20	.0	.141	.132	.129	.121	.121	.112	.115	.106
		.4	.098	.102	.089	.094	.085	.087	.082	.084
		.8	.069	.071	.061	.066	.059	.060	.058	.059
	60	.0	.042	.039	.041	.038	.040	.037	.039	.037
		.4	.029	.031	.028	.030	.028	.030	.027	.029
		.8	.018	.021	.018	.021	.017	.020	.017	.020
g-h-1	20	.0	.218	.245	.192	.215	.195	.222	.179	.202
		.4	.184	.239	.159	.207	.163	.214	.148	.194
		.8	.157	.240	.133	.207	.139	.210	.125	.190
	60	.0	.071	.080	.068	.076	.071	.081	.069	.079
		.4	.061	.081	.058	.077	.063	.082	.061	.080
		.8	.060	.081	.056	.076	.063	.082	.060	.079
g-h-2	20	.0	.197	.109	.176	.097	.177	.096	.165	.090
		.4	.163	.092	.144	.080	.145	.081	.136	.076
		.8	.136	.085	.118	.072	.116	.077	.109	.071
	60	.0	.058	.031	.056	.030	.056	.030	.055	.030
		.4	.047	.027	.045	.026	.045	.027	.044	.026
		.8	.036	.024	.035	.023	.035	.024	.034	.024
g-h-3	20	.0	.256	.215	.224	.186	.227	.187	.206	.169
		.4	.237	.221	.204	.189	.208	.188	.186	.167
		.8	.249	.228	.210	.193	.213	.191	.187	.169
	60	.0	.087	.072	.083	.068	.088	.073	.085	.070
		.4	.088	.078	.083	.073	.091	.078	.088	.075
		.8	.100	.084	.094	.079	.103	.085	.099	.081
g-h-4	20	.0	.150	.091	.136	.084	.130	.078	.123	.075
		.4	.117	.068	.106	.062	.101	.059	.097	.058
		.8	.093	.044	.082	.041	.077	.039	.074	.040
	60	.0	.044	.027	.043	.027	.042	.026	.041	.026
		.4	.033	.020	.032	.020	.031	.019	.031	.019
		.8	.023	.012	.022	.012	.022	.012	.021	.012
Hyper	20	.0	.177	.128	.162	.118	.160	.114	.151	.109
		.4	.111	.098	.101	.090	.099	.087	.094	.084
		.8	.032	.072	.029	.067	.028	.063	.027	.062
	60	.0	.055	.039	.053	.038	.053	.037	.052	.037
		.4	.033	.030	.032	.029	.032	.028	.031	.028
		.8	.017	.022	.016	.021	.016	.020	.016	.020

Anmerkung: Vorpalte 1 gibt die Verteilung an, Vorpalte 2 den Stichprobenumfang N , Vorpalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-3. $\overline{B_{MC}^2}$

			d'_s		\hat{d}'_{c1}		\hat{d}'_{c2}		d'^{Boot}		\hat{d}'^{Boot}_{c1}		\hat{d}'^{Boot}_{c2}		
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	
Normal	20	.0	.000	.001	.000	.000	.000	.000	.001	.001	.004	.004	.002	.003	
		.4	.001	.001	.000	.000	.000	.000	.001	.001	.004	.004	.002	.004	
		.8	.001	.002	.000	.000	.000	.000	.001	.001	.004	.004	.003	.004	
	60	.0	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	
		.4	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	
		.8	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	
Gleich	20	.0	.000	.000	.001	.000	.000	.000	.001	.001	.004	.004	.002	.003	
		.4	.000	.000	.001	.001	.000	.000	.001	.001	.004	.005	.003	.004	
		.8	.000	.000	.001	.001	.000	.000	.001	.001	.005	.005	.004	.004	
	60	.0	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	
		.4	.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.000	.000	
		.8	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	
g-h-1	20	.0	.013	.024	.005	.012	.008	.014	.001	.002	.000	.000	.000	.000	
		.4	.016	.023	.007	.011	.010	.013	.001	.002	.000	.000	.000	.000	
		.8	.024	.030	.012	.016	.014	.017	.002	.003	.000	.000	.000	.000	
	60	.0	.003	.005	.001	.003	.002	.004	.000	.001	.000	.000	.000	.000	
		.4	.003	.005	.002	.003	.002	.003	.000	.001	.000	.000	.000	.000	
		.8	.005	.006	.003	.004	.004	.004	.001	.001	.000	.000	.000	.000	
g-h-2	20	.0	.004	.004	.001	.001	.002	.002	.000	.000	.003	.002	.001	.001	
		.4	.006	.005	.001	.002	.003	.002	.000	.000	.003	.002	.001	.001	
		.8	.013	.009	.005	.003	.006	.003	.000	.000	.001	.001	.001	.001	
	60	.0	.001	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	
		.4	.001	.001	.000	.000	.001	.000	.000	.000	.000	.000	.000	.000	
		.8	.001	.001	.001	.000	.001	.000	.000	.000	.000	.000	.000	.000	
g-h-3	20	.0	.025	.029	.013	.016	.017	.018	.004	.005	.000	.001	.001	.001	
		.4	.031	.032	.017	.018	.021	.020	.005	.005	.001	.001	.002	.001	
		.8	.045	.041	.027	.024	.030	.025	.007	.007	.002	.002	.003	.002	
	60	.0	.007	.008	.005	.006	.006	.006	.001	.002	.001	.001	.001	.001	
		.4	.008	.009	.005	.006	.006	.007	.002	.002	.001	.001	.001	.001	
		.8	.011	.011	.009	.008	.009	.008	.002	.003	.001	.001	.002	.002	
g-h-4	20	.0	.001	.000	.000	.001	.000	.001	.001	.001	.004	.004	.002	.003	
		.4	.001	.000	.000	.001	.000	.001	.001	.001	.001	.005	.004	.003	.004
		.8	.001	.001	.000	.001	.000	.001	.001	.001	.001	.005	.005	.004	.004
	60	.0	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	
		.4	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	
		.8	.000	.000	.000	.000	.000	.000	.000	.000	.001	.000	.000	.000	
Hyper	20	.0	.000	.000	.001	.001	.000	.000	.000	.001	.004	.004	.002	.003	
		.4	.001	.000	.000	.001	.000	.000	.000	.001	.003	.004	.002	.004	
		.8	.001	.000	.000	.001	.000	.001	.001	.001	.001	.004	.004	.003	.004
	60	.0	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	
		.4	.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.000	.000	
		.8	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-4. \overline{MSE}_{MC}

		d'_s		\hat{d}'_{c1}		\hat{d}'_{c2}		d'^{Boot}_s		\hat{d}'^{Boot}_{c1}		\hat{d}'^{Boot}_{c2}		
		1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	
Normal	20	.0	.121	.137	.111	.125	.115	.128	.111	.122	.105	.116	.108	.118
		.4	.079	.109	.073	.099	.075	.101	.073	.097	.070	.093	.071	.094
		.8	.046	.086	.042	.078	.043	.079	.042	.076	.042	.073	.042	.074
	60	.0	.039	.042	.038	.040	.038	.041	.038	.040	.037	.039	.037	.040
		.4	.025	.034	.024	.033	.025	.033	.024	.033	.024	.032	.024	.032
		.8	.014	.026	.013	.025	.013	.025	.013	.025	.013	.025	.013	.025
Gleich	20	.0	.115	.124	.106	.114	.110	.116	.104	.107	.099	.102	.101	.103
		.4	.070	.094	.066	.087	.067	.089	.064	.082	.063	.079	.063	.079
		.8	.031	.070	.029	.065	.030	.065	.028	.060	.030	.059	.030	.059
	60	.0	.037	.037	.036	.036	.036	.036	.035	.036	.035	.035	.035	.035
		.4	.022	.029	.022	.029	.022	.029	.022	.028	.021	.028	.022	.028
		.8	.009	.021	.009	.021	.009	.021	.009	.020	.009	.020	.009	.020
g-h-1	20	.0	.181	.226	.159	.199	.168	.205	.170	.211	.156	.192	.163	.198
		.4	.151	.199	.131	.174	.139	.179	.141	.184	.129	.168	.135	.172
		.8	.144	.195	.123	.169	.128	.171	.130	.176	.118	.159	.122	.161
	60	.0	.064	.077	.061	.073	.062	.074	.066	.079	.064	.076	.065	.077
		.4	.053	.069	.050	.066	.051	.066	.055	.071	.053	.069	.054	.069
		.8	.051	.066	.048	.063	.048	.063	.053	.068	.051	.066	.051	.066
g-h-2	20	.0	.162	.098	.145	.088	.152	.089	.153	.088	.144	.083	.148	.084
		.4	.132	.081	.118	.072	.123	.073	.125	.074	.118	.070	.121	.071
		.8	.127	.077	.110	.066	.114	.067	.117	.070	.109	.066	.112	.066
	60	.0	.051	.029	.050	.028	.050	.028	.051	.028	.050	.028	.050	.028
		.4	.041	.025	.039	.024	.040	.024	.040	.025	.039	.024	.040	.024
		.8	.035	.022	.034	.021	.034	.021	.035	.023	.034	.022	.035	.022
g-h-3	20	.0	.215	.198	.188	.171	.199	.177	.202	.177	.183	.160	.192	.164
		.4	.186	.179	.159	.154	.170	.159	.172	.159	.154	.143	.162	.147
		.8	.196	.181	.166	.153	.173	.155	.174	.155	.156	.138	.160	.140
	60	.0	.079	.068	.075	.065	.077	.066	.082	.070	.079	.067	.081	.068
		.4	.071	.064	.067	.060	.068	.061	.075	.065	.072	.063	.073	.063
		.8	.072	.066	.068	.062	.068	.062	.075	.068	.072	.065	.072	.065
g-h-4	20	.0	.123	.084	.113	.078	.117	.079	.112	.073	.107	.071	.109	.071
		.4	.085	.063	.078	.059	.081	.059	.077	.056	.075	.055	.075	.055
		.8	.055	.044	.050	.041	.051	.041	.048	.039	.048	.040	.048	.040
	60	.0	.039	.026	.038	.025	.038	.025	.038	.025	.037	.024	.037	.024
		.4	.025	.019	.025	.019	.025	.019	.025	.018	.024	.018	.024	.018
		.8	.014	.013	.014	.012	.014	.012	.014	.012	.014	.012	.014	.012
Hyper	20	.0	.152	.120	.140	.111	.144	.113	.142	.109	.134	.104	.137	.104
		.4	.108	.093	.099	.086	.102	.087	.100	.084	.095	.081	.097	.081
		.8	.077	.069	.070	.064	.072	.065	.070	.062	.067	.061	.068	.061
	60	.0	.049	.037	.048	.036	.049	.036	.048	.035	.047	.035	.048	.035
		.4	.033	.029	.032	.028	.033	.028	.033	.027	.032	.027	.032	.027
		.8	.022	.021	.021	.021	.021	.021	.021	.020	.021	.020	.021	.020

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-5. B_{MC}^2

			$d_s^{S_1}$		\hat{d}^{S_1}		$d_s^{S_1,Boot}$		$\hat{d}^{S_1,Boot}$	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	.002	.002	.000	.000	.001	.001	.004	.004
		.4	.002	.002	.000	.000	.001	.001	.004	.004
		.8	.002	.002	.000	.000	.001	.001	.004	.004
	60	.0	.000	.000	.000	.000	.000	.000	.000	.000
		.4	.000	.000	.000	.000	.000	.000	.000	.000
		.8	.000	.000	.000	.000	.000	.000	.000	.000
Gleich	20	.0	.000	.001	.001	.000	.001	.001	.005	.004
		.4	.000	.000	.001	.000	.001	.001	.005	.005
		.8	.000	.000	.001	.000	.001	.001	.005	.005
	60	.0	.000	.000	.000	.000	.000	.000	.000	.000
		.4	.000	.000	.000	.000	.000	.000	.000	.001
		.8	.000	.000	.000	.000	.000	.000	.000	.000
g-h-1	20	.0	.033	.035	.018	.020	.003	.003	.000	.000
		.4	.034	.032	.019	.017	.003	.003	.000	.000
		.8	.033	.033	.018	.019	.003	.003	.000	.000
	60	.0	.007	.007	.005	.005	.001	.001	.000	.000
		.4	.007	.007	.005	.005	.001	.001	.000	.000
		.8	.007	.007	.005	.005	.001	.001	.000	.000
g-h-2	20	.0	.008	.007	.002	.003	.000	.000	.002	.002
		.4	.011	.008	.004	.003	.000	.000	.002	.002
		.8	.016	.010	.007	.004	.000	.000	.001	.001
	60	.0	.001	.001	.000	.000	.000	.000	.000	.000
		.4	.002	.001	.001	.000	.000	.000	.000	.000
		.8	.002	.001	.001	.001	.000	.000	.000	.000
g-h-3	20	.0	.045	.041	.027	.024	.008	.007	.002	.002
		.4	.052	.043	.032	.025	.009	.007	.003	.002
		.8	.056	.045	.035	.027	.009	.007	.003	.002
	60	.0	.013	.011	.010	.008	.004	.002	.002	.001
		.4	.013	.011	.010	.009	.003	.003	.002	.002
		.8	.014	.012	.011	.009	.003	.003	.002	.002
g-h-4	20	.0	.001	.001	.001	.001	.001	.001	.005	.004
		.4	.001	.001	.000	.001	.001	.001	.005	.005
		.8	.002	.001	.000	.001	.001	.001	.005	.005
	60	.0	.000	.000	.000	.000	.000	.000	.000	.000
		.4	.000	.000	.000	.000	.000	.000	.000	.000
		.8	.000	.000	.000	.000	.000	.000	.001	.000
Hyper	20	.0	.000	.000	.001	.001	.001	.001	.004	.005
		.4	.001	.000	.000	.001	.001	.001	.004	.005
		.8	.001	.000	.000	.001	.001	.001	.004	.004
	60	.0	.000	.000	.000	.000	.000	.000	.000	.000
		.4	.000	.000	.000	.000	.000	.000	.000	.001
		.8	.000	.000	.000	.000	.000	.000	.000	.000

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-6. MSE_{MC}

			$d_s^{S_1}$		\hat{d}^{S_1}		$d_s^{S_1,Boot}$		$\hat{d}^{S_1,Boot}$	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	.144	.148	.131	.134	.125	.128	.119	.122
		.4	.098	.117	.089	.106	.085	.102	.082	.097
		.8	.054	.089	.048	.081	.047	.078	.047	.075
	60	.0	.044	.044	.042	.042	.042	.042	.041	.041
		.4	.030	.036	.029	.035	.028	.034	.028	.034
		.8	.015	.027	.015	.026	.015	.026	.015	.025
Gleich	20	.0	.124	.129	.115	.118	.108	.108	.104	.103
		.4	.078	.098	.072	.091	.069	.083	.067	.080
		.8	.034	.072	.032	.066	.030	.061	.032	.060
	60	.0	.039	.038	.038	.037	.037	.036	.037	.036
		.4	.024	.030	.023	.029	.023	.029	.023	.028
		.8	.010	.021	.010	.021	.010	.020	.010	.020
g-h-1	20	.0	.279	.273	.246	.239	.241	.244	.220	.222
		.4	.228	.234	.198	.204	.197	.209	.179	.191
		.8	.177	.209	.151	.180	.154	.185	.139	.168
	60	.0	.090	.088	.086	.084	.090	.089	.087	.086
		.4	.075	.078	.071	.074	.075	.079	.073	.077
		.8	.060	.070	.056	.066	.061	.071	.059	.069
g-h-2	20	.0	.143	.108	.127	.096	.124	.094	.116	.088
		.4	.129	.091	.113	.079	.116	.081	.109	.076
		.8	.131	.082	.113	.070	.119	.074	.111	.069
	60	.0	.042	.030	.040	.029	.040	.030	.040	.029
		.4	.037	.026	.035	.025	.036	.026	.035	.026
		.8	.034	.023	.032	.022	.034	.024	.033	.023
g-h-3	20	.0	.277	.235	.241	.203	.231	.201	.208	.180
		.4	.243	.209	.208	.179	.204	.178	.183	.160
		.8	.222	.193	.189	.163	.189	.163	.168	.145
	60	.0	.091	.078	.086	.073	.089	.078	.085	.075
		.4	.081	.072	.076	.068	.081	.072	.077	.069
		.8	.077	.070	.072	.065	.078	.071	.074	.068
g-h-4	20	.0	.100	.085	.092	.078	.086	.071	.083	.069
		.4	.073	.064	.066	.060	.064	.055	.063	.055
		.8	.054	.045	.048	.041	.046	.039	.047	.040
	60	.0	.030	.025	.029	.024	.028	.024	.028	.024
		.4	.020	.019	.020	.018	.020	.018	.019	.018
		.8	.013	.013	.013	.012	.012	.012	.012	.012
Hyper	20	.0	.130	.123	.120	.113	.116	.108	.111	.103
		.4	.092	.095	.084	.088	.082	.084	.079	.081
		.8	.063	.071	.057	.065	.055	.063	.055	.061
	60	.0	.040	.037	.039	.036	.039	.035	.038	.035
		.4	.027	.029	.027	.028	.026	.028	.026	.027
		.8	.018	.021	.017	.021	.017	.020	.017	.020

Anmerkung: Vorpalte 1 gibt die Verteilung an, Vorpalte 2 den Stichprobenumfang N , Vorpalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-7. \overline{B}_{MC}^2

			$d_s^{S_2}$		\hat{d}^{S_2}		$d_s^{S_2,Boot}$		$\hat{d}^{S_2,Boot}$	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	.002	.002	.000	.000	.001	.001	.004	.004
		.4	.002	.001	.000	.000	.001	.001	.004	.005
		.8	.002	.002	.000	.000	.001	.001	.004	.004
	60	.0	.000	.000	.000	.000	.000	.000	.000	.000
		.4	.000	.000	.000	.000	.000	.000	.000	.000
		.8	.000	.000	.000	.000	.000	.000	.000	.000
Gleich	20	.0	.000	.000	.000	.000	.001	.001	.005	.004
		.4	.000	.000	.001	.001	.001	.001	.005	.005
		.8	.000	.000	.001	.000	.001	.001	.005	.005
	60	.0	.000	.000	.000	.000	.000	.000	.000	.000
		.4	.000	.000	.000	.000	.000	.000	.000	.000
		.8	.000	.000	.000	.000	.000	.000	.000	.000
g-h-1	20	.0	.036	.034	.021	.019	.004	.003	.000	.000
		.4	.034	.031	.019	.017	.003	.002	.000	.000
		.8	.033	.034	.018	.019	.003	.003	.000	.000
	60	.0	.007	.008	.005	.005	.001	.001	.000	.000
		.4	.007	.007	.005	.005	.001	.001	.000	.000
		.8	.007	.007	.005	.005	.001	.001	.000	.000
g-h-2	20	.0	.036	.024	.021	.012	.000	.000	.001	.001
		.4	.027	.018	.014	.008	.000	.000	.001	.001
		.8	.022	.012	.010	.005	.000	.000	.001	.001
	60	.0	.004	.003	.002	.002	.000	.000	.000	.000
		.4	.004	.002	.002	.001	.000	.000	.000	.000
		.8	.002	.002	.001	.001	.000	.000	.000	.000
g-h-3	20	.0	.078	.066	.053	.043	.014	.011	.006	.004
		.4	.072	.059	.047	.038	.013	.010	.005	.003
		.8	.062	.052	.040	.032	.010	.009	.003	.003
	60	.0	.018	.017	.014	.014	.004	.004	.002	.003
		.4	.018	.016	.014	.012	.004	.004	.003	.002
		.8	.016	.014	.013	.011	.004	.004	.002	.002
g-h-4	20	.0	.006	.004	.001	.000	.001	.001	.006	.005
		.4	.004	.002	.001	.000	.002	.001	.006	.006
		.8	.003	.001	.000	.000	.001	.001	.006	.005
	60	.0	.001	.000	.000	.000	.000	.000	.000	.000
		.4	.000	.000	.000	.000	.000	.000	.000	.001
		.8	.000	.000	.000	.000	.000	.000	.000	.000
Hyper	20	.0	.010	.003	.003	.000	.000	.000	.003	.003
		.4	.012	.003	.005	.000	.000	.000	.003	.002
		.8	.009	.002	.003	.000	.001	.000	.004	.003
	60	.0	.001	.000	.001	.000	.000	.000	.000	.000
		.4	.001	.000	.001	.000	.000	.000	.000	.000
		.8	.001	.000	.000	.000	.000	.000	.000	.000

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-8. MSE_{MC}

			$d_s^{S_2}$		\hat{d}^{S_2}		$d_s^{S_2,Boot}$		$\hat{d}^{S_2,Boot}$	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	.147	.146	.134	.133	.129	.126	.122	.120
		.4	.099	.116	.090	.106	.086	.102	.083	.098
		.8	.055	.090	.049	.081	.048	.078	.048	.076
	60	.0	.044	.044	.042	.042	.042	.042	.041	.041
		.4	.029	.036	.028	.035	.028	.034	.028	.034
		.8	.016	.027	.015	.026	.015	.026	.015	.025
Gleich	20	.0	.127	.122	.117	.112	.110	.109	.105	.104
		.4	.079	.093	.073	.086	.070	.084	.068	.081
		.8	.035	.070	.032	.064	.031	.061	.032	.060
	60	.0	.039	.038	.038	.037	.037	.037	.037	.036
		.4	.024	.030	.023	.029	.023	.029	.023	.029
		.8	.010	.021	.010	.021	.010	.020	.010	.020
g-h-1	20	.0	.290	.308	.254	.271	.251	.259	.228	.236
		.4	.230	.258	.200	.226	.200	.223	.181	.203
		.8	.177	.222	.151	.192	.154	.195	.139	.177
	60	.0	.091	.095	.086	.091	.090	.094	.087	.090
		.4	.076	.084	.072	.080	.077	.083	.074	.081
		.8	.060	.072	.057	.068	.062	.072	.060	.070
g-h-2	20	.0	.428	.340	.382	.303	.359	.287	.331	.266
		.4	.303	.235	.269	.208	.263	.208	.244	.192
		.8	.189	.129	.165	.112	.168	.117	.156	.109
	60	.0	.107	.088	.103	.085	.100	.083	.098	.081
		.4	.079	.065	.075	.062	.075	.062	.073	.061
		.8	.051	.037	.048	.035	.049	.036	.048	.035
g-h-3	20	.0	.448	.416	.394	.366	.387	.349	.349	.315
		.4	.345	.332	.300	.290	.300	.285	.269	.257
		.8	.253	.232	.216	.198	.218	.197	.195	.176
	60	.0	.144	.138	.137	.132	.144	.136	.139	.131
		.4	.120	.112	.114	.106	.123	.112	.118	.108
		.8	.092	.083	.087	.078	.095	.084	.091	.081
g-h-4	20	.0	.237	.191	.214	.173	.184	.154	.174	.147
		.4	.162	.125	.146	.113	.128	.106	.123	.102
		.8	.082	.066	.073	.060	.066	.058	.065	.058
	60	.0	.061	.052	.059	.050	.056	.049	.055	.048
		.4	.039	.035	.038	.034	.036	.033	.036	.033
		.8	.020	.018	.019	.018	.018	.017	.018	.017
Hyper	20	.0	.339	.163	.306	.148	.309	.152	.287	.143
		.4	.275	.127	.247	.115	.247	.118	.230	.111
		.8	.184	.092	.164	.083	.162	.086	.152	.082
	60	.0	.080	.050	.077	.048	.073	.048	.071	.047
		.4	.059	.037	.057	.036	.053	.036	.052	.035
		.8	.037	.027	.036	.027	.033	.026	.032	.026

Anmerkung: Vorpalte 1 gibt die Verteilung an, Vorpalte 2 den Stichprobenumfang N , Vorpalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-9. $\overline{B_{MC}^2}$

			$d_{D,s,R}$		$\hat{d}_{D,R}$		$d_{D,s,R}^{Boot}$		$\hat{d}_{D,R}^{Boot}$	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	.007	.011	.000	.001	.002	.002	.001	.001
		.4	.005	.011	.000	.001	.001	.001	.002	.001
		.8	.002	.013	.001	.001	.000	.001	.004	.002
	60	.0	.001	.001	.000	.000	.000	.000	.000	.000
		.4	.001	.001	.000	.000	.000	.000	.000	.000
		.8	.000	.001	.000	.000	.000	.000	.000	.000
Gleich	20	.0	.010	.014	.001	.002	.001	.000	.002	.002
		.4	.008	.013	.000	.001	.000	.000	.004	.003
		.8	.007	.013	.000	.002	.000	.000	.005	.004
	60	.0	.001	.001	.000	.000	.000	.000	.000	.000
		.4	.001	.001	.000	.000	.000	.000	.000	.000
		.8	.001	.001	.000	.000	.000	.000	.000	.000
g-h-1	20	.0	.004	.009	.000	.000	.003	.003	.000	.000
		.4	.002	.008	.001	.000	.002	.002	.001	.001
		.8	.000	.012	.003	.001	.000	.002	.003	.001
	60	.0	.000	.001	.000	.000	.000	.000	.000	.000
		.4	.000	.001	.000	.000	.000	.000	.000	.000
		.8	.000	.001	.000	.000	.000	.000	.000	.000
g-h-2	20	.0	.007	.007	.000	.001	.001	.001	.001	.001
		.4	.005	.009	.000	.001	.000	.001	.002	.001
		.8	.004	.014	.000	.003	.000	.001	.004	.001
	60	.0	.001	.001	.000	.000	.000	.000	.000	.000
		.4	.001	.001	.000	.000	.000	.000	.000	.000
		.8	.000	.001	.000	.000	.000	.000	.000	.000
g-h-3	20	.0	.005	.007	.000	.000	.003	.003	.000	.000
		.4	.002	.008	.001	.000	.002	.002	.001	.000
		.8	.000	.011	.003	.001	.000	.002	.003	.001
	60	.0	.000	.001	.000	.000	.001	.000	.000	.000
		.4	.000	.001	.000	.000	.000	.000	.000	.000
		.8	.000	.001	.000	.000	.000	.000	.000	.000
g-h-4	20	.0	.011	.009	.001	.001	.001	.001	.002	.001
		.4	.010	.010	.001	.002	.000	.001	.002	.001
		.8	.008	.013	.000	.003	.000	.000	.005	.002
	60	.0	.001	.001	.000	.000	.000	.000	.000	.000
		.4	.001	.001	.000	.000	.000	.000	.000	.000
		.8	.001	.001	.000	.000	.000	.000	.000	.000
Hyper	20	.0	.003	.004	.008	.002	.000	.004	.008	.016
		.4	.002	.004	.007	.002	.000	.005	.010	.018
		.8	.008	.004	.004	.005	.003	.013	.018	.030
	60	.0	.001	.002	.002	.003	.001	.004	.002	.006
		.4	.001	.003	.004	.004	.001	.005	.003	.007
		.8	.001	.003	.002	.006	.001	.006	.003	.009

Anmerkung: Vorpalte 1 gibt die Verteilung an, Vorpalte 2 den Stichprobenumfang N , Vorpalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-10. MSE_{MC}

			$d_{D,s,R}$		$\hat{d}_{D,R}$		$d_{D,s,R}^{Boot}$		$\hat{d}_{D,R}^{Boot}$	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	.204	.231	.171	.191	.209	.229	.180	.198
		.4	.142	.208	.119	.172	.155	.217	.136	.188
		.8	.090	.190	.077	.154	.118	.216	.106	.187
	60	.0	.057	.060	.054	.057	.058	.062	.055	.059
		.4	.040	.053	.038	.050	.042	.055	.040	.053
		.8	.024	.047	.023	.044	.028	.051	.026	.048
Gleich	20	.0	.290	.317	.242	.264	.266	.277	.232	.242
		.4	.191	.272	.158	.225	.184	.247	.163	.216
		.8	.115	.222	.094	.182	.124	.215	.112	.190
	60	.0	.079	.079	.075	.075	.077	.077	.074	.074
		.4	.053	.067	.050	.063	.052	.066	.050	.063
		.8	.029	.052	.027	.049	.030	.052	.029	.050
g-h-1	20	.0	.130	.160	.109	.131	.153	.180	.130	.154
		.4	.098	.154	.083	.127	.128	.187	.110	.161
		.8	.072	.171	.065	.138	.114	.226	.102	.195
	60	.0	.037	.041	.035	.039	.041	.046	.038	.043
		.4	.027	.040	.026	.037	.032	.046	.030	.043
		.8	.019	.041	.019	.038	.025	.049	.024	.047
g-h-2	20	.0	.185	.120	.154	.099	.203	.127	.176	.109
		.4	.146	.113	.122	.091	.172	.128	.151	.110
		.8	.127	.117	.106	.092	.172	.144	.153	.124
	60	.0	.051	.031	.048	.029	.055	.033	.053	.032
		.4	.040	.027	.038	.025	.045	.030	.043	.029
		.8	.033	.027	.031	.025	.040	.032	.038	.031
g-h-3	20	.0	.126	.123	.105	.101	.151	.143	.128	.122
		.4	.096	.123	.082	.099	.128	.151	.110	.129
		.8	.075	.138	.068	.110	.121	.186	.108	.160
	60	.0	.035	.032	.033	.030	.039	.036	.037	.034
		.4	.026	.030	.025	.028	.031	.036	.029	.034
		.8	.020	.032	.019	.030	.026	.040	.025	.038
g-h-4	20	.0	.250	.169	.208	.140	.253	.157	.220	.136
		.4	.195	.148	.161	.121	.207	.145	.181	.127
		.8	.148	.132	.122	.105	.181	.145	.161	.127
	60	.0	.068	.044	.064	.041	.070	.044	.067	.042
		.4	.052	.035	.049	.033	.055	.037	.052	.035
		.8	.037	.029	.035	.027	.042	.033	.040	.032
Hyper	20	.0	.284	.445	.251	.383	.349	.414	.309	.371
		.4	.199	.409	.178	.353	.267	.410	.241	.369
		.8	.226	.371	.193	.322	.265	.383	.245	.351
	60	.0	.086	.096	.084	.093	.105	.094	.102	.092
		.4	.056	.082	.056	.081	.072	.083	.071	.083
		.8	.067	.069	.066	.068	.113	.075	.111	.075

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-11. B_{MC}^2

			$d'_{s,R}$		$\hat{d}'_{c1,R}$		$\hat{d}'_{c2,R}$		$d'^{Boot}_{s,R}$		$\hat{d}'^{Boot}_{c1,R}$		$\hat{d}'^{Boot}_{c2,R}$	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	.004	.010	.000	.001	.000	.001	.002	.003	.001	.001	.000	.000
		.4	.004	.009	.000	.000	.000	.001	.002	.002	.001	.001	.000	.000
		.8	.007	.010	.000	.001	.001	.001	.002	.001	.001	.001	.000	.001
	60	.0	.000	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
		.4	.000	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
		.8	.001	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
Gleich	20	.0	.006	.012	.000	.001	.001	.002	.001	.001	.002	.002	.000	.001
		.4	.006	.010	.000	.001	.001	.002	.001	.000	.002	.002	.000	.001
		.8	.008	.011	.000	.001	.001	.001	.000	.000	.002	.003	.001	.002
	60	.0	.001	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
		.4	.001	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
		.8	.001	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
g-h-1	20	.0	.002	.007	.001	.000	.000	.001	.003	.004	.000	.000	.000	.000
		.4	.003	.006	.000	.000	.000	.000	.003	.002	.000	.001	.000	.000
		.8	.005	.008	.000	.000	.000	.001	.003	.002	.000	.001	.000	.000
	60	.0	.000	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
		.4	.000	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
		.8	.000	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
g-h-2	20	.0	.004	.006	.000	.000	.001	.001	.002	.002	.001	.000	.000	.000
		.4	.006	.007	.000	.001	.001	.001	.002	.002	.001	.000	.000	.000
		.8	.012	.010	.001	.001	.003	.002	.002	.001	.001	.001	.000	.000
	60	.0	.000	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
		.4	.001	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
		.8	.001	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
g-h-3	20	.0	.002	.006	.001	.000	.000	.001	.003	.003	.000	.000	.000	.000
		.4	.003	.006	.000	.000	.000	.000	.004	.003	.000	.000	.000	.000
		.8	.007	.008	.000	.000	.001	.001	.003	.002	.000	.000	.000	.000
	60	.0	.000	.000	.000	.000	.000	.000	.001	.000	.000	.000	.000	.000
		.4	.000	.001	.000	.000	.000	.000	.001	.000	.000	.000	.000	.000
		.8	.001	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
g-h-4	20	.0	.007	.008	.000	.001	.002	.002	.001	.001	.001	.001	.000	.000
		.4	.010	.008	.001	.001	.003	.002	.001	.001	.001	.001	.000	.001
		.8	.016	.010	.003	.002	.005	.002	.001	.000	.002	.002	.001	.001
	60	.0	.001	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
		.4	.001	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
		.8	.001	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
Hyper	20	.0	.002	.004	.010	.002	.006	.002	.000	.003	.007	.014	.003	.012
		.4	.002	.004	.008	.002	.005	.002	.000	.002	.006	.013	.003	.011
		.8	.002	.005	.006	.002	.004	.002	.000	.003	.005	.014	.003	.013
	60	.0	.001	.002	.003	.003	.002	.003	.001	.003	.002	.006	.001	.005
		.4	.001	.002	.003	.004	.003	.003	.001	.004	.003	.006	.002	.006
		.8	.001	.002	.002	.003	.002	.003	.001	.003	.002	.006	.002	.006

Anmerkung: Vorpalte 1 gibt die Verteilung an, Vorpalte 2 den Stichprobenumfang N , Vorpalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-12. \overline{MSE}_{MC}

		$d'_{s,R}$		$\hat{d}'_{c1,R}$		$\hat{d}'_{c2,R}$		$d'^{Boot}_{s,R}$		$\hat{d}'^{Boot}_{c1,R}$		$\hat{d}'^{Boot}_{c2,R}$		
		1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	
Normal	20	.0	.172	.217	.146	.180	.156	.189	.185	.224	.159	.192	.171	.202
		.4	.126	.184	.106	.152	.113	.160	.140	.194	.121	.167	.129	.175
		.8	.089	.155	.072	.126	.076	.130	.099	.162	.085	.140	.090	.144
	60	.0	.051	.057	.048	.054	.049	.055	.053	.060	.051	.057	.052	.058
		.4	.036	.049	.034	.046	.035	.047	.038	.052	.036	.049	.037	.050
		.8	.024	.041	.022	.038	.022	.039	.025	.043	.024	.041	.024	.041
Gleich	20	.0	.252	.302	.213	.252	.228	.264	.243	.275	.211	.239	.226	.249
		.4	.168	.245	.140	.204	.150	.213	.167	.227	.146	.199	.154	.206
		.8	.097	.191	.077	.157	.083	.161	.098	.174	.087	.153	.091	.157
	60	.0	.073	.076	.069	.072	.071	.073	.072	.075	.069	.072	.071	.072
		.4	.047	.064	.045	.060	.046	.061	.048	.063	.046	.060	.047	.061
		.8	.025	.048	.023	.045	.023	.046	.025	.048	.024	.046	.024	.046
g-h-1	20	.0	.106	.146	.091	.120	.096	.127	.130	.173	.111	.147	.119	.156
		.4	.089	.130	.075	.107	.080	.113	.113	.157	.096	.134	.103	.141
		.8	.083	.127	.067	.103	.072	.107	.101	.149	.085	.128	.091	.132
	60	.0	.031	.039	.030	.036	.031	.037	.035	.044	.033	.041	.034	.042
		.4	.025	.035	.023	.033	.024	.034	.028	.040	.027	.038	.027	.039
		.8	.022	.033	.021	.031	.021	.031	.025	.038	.024	.036	.024	.036
g-h-2	20	.0	.156	.111	.131	.091	.140	.095	.179	.121	.155	.104	.166	.108
		.4	.135	.099	.112	.080	.120	.084	.157	.111	.136	.095	.145	.100
		.8	.134	.093	.106	.073	.113	.075	.147	.102	.127	.088	.133	.090
	60	.0	.046	.029	.043	.027	.044	.028	.050	.032	.048	.030	.049	.031
		.4	.038	.025	.035	.023	.036	.024	.042	.028	.040	.027	.040	.027
		.8	.034	.022	.031	.021	.032	.021	.037	.025	.035	.024	.036	.024
g-h-3	20	.0	.102	.112	.087	.092	.092	.097	.128	.136	.108	.115	.117	.122
		.4	.087	.102	.072	.083	.077	.088	.111	.125	.093	.106	.101	.112
		.8	.091	.101	.073	.081	.078	.084	.111	.120	.093	.102	.099	.105
	60	.0	.029	.030	.028	.028	.029	.028	.034	.034	.032	.032	.033	.033
		.4	.024	.027	.023	.025	.023	.026	.028	.031	.026	.029	.027	.030
		.8	.023	.026	.022	.024	.022	.024	.026	.030	.025	.028	.025	.028
g-h-4	20	.0	.218	.158	.183	.131	.195	.137	.231	.152	.200	.132	.213	.136
		.4	.182	.135	.150	.111	.160	.115	.193	.132	.168	.115	.178	.119
		.8	.159	.108	.126	.086	.135	.089	.160	.106	.140	.093	.146	.095
	60	.0	.063	.042	.059	.040	.061	.040	.065	.043	.062	.041	.064	.041
		.4	.048	.034	.045	.032	.046	.032	.050	.035	.048	.034	.049	.034
		.8	.036	.026	.033	.024	.034	.024	.038	.027	.036	.026	.036	.026
Hyper	20	.0	.252	.432	.226	.373	.234	.385	.322	.410	.285	.366	.297	.378
		.4	.192	.379	.173	.327	.179	.337	.260	.379	.231	.339	.240	.351
		.8	.157	.306	.140	.263	.144	.268	.232	.298	.206	.269	.211	.274
	60	.0	.081	.094	.079	.092	.079	.092	.099	.093	.097	.091	.097	.091
		.4	.058	.080	.058	.078	.058	.079	.074	.081	.073	.080	.073	.080
		.8	.046	.064	.046	.062	.046	.063	.065	.065	.064	.064	.064	.064

Anmerkung: Vorderspalte 1 gibt die Verteilung an, Vorderspalte 2 den Stichprobenumfang N , Vorderspalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-13. $\overline{B_{MC}^2}$

			$d_{s,R}^{S_1}$		$\hat{d}_R^{S_1}$		$d_{s,R}^{S_1,Boot}$		$\hat{d}_R^{S_1,Boot}$	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	.012	.014	.001	.002	.001	.001	.002	.002
		.4	.013	.013	.001	.001	.001	.001	.002	.002
		.8	.013	.013	.001	.001	.001	.001	.002	.002
	60	.0	.001	.001	.000	.000	.000	.000	.000	.000
		.4	.001	.001	.000	.000	.000	.000	.000	.000
		.8	.001	.001	.000	.000	.000	.000	.000	.000
Gleich	20	.0	.014	.016	.002	.002	.000	.000	.004	.003
		.4	.013	.014	.001	.002	.000	.000	.004	.004
		.8	.013	.013	.002	.001	.000	.000	.004	.004
	60	.0	.001	.001	.000	.000	.000	.000	.000	.000
		.4	.001	.001	.000	.000	.000	.000	.000	.000
		.8	.001	.001	.000	.000	.000	.000	.000	.000
g-h-1	20	.0	.011	.012	.001	.001	.001	.002	.001	.001
		.4	.011	.010	.001	.001	.001	.001	.001	.002
		.8	.011	.011	.001	.001	.001	.001	.001	.001
	60	.0	.001	.001	.000	.000	.000	.000	.000	.000
		.4	.001	.001	.000	.000	.000	.000	.000	.000
		.8	.001	.001	.000	.000	.000	.000	.000	.000
g-h-2	20	.0	.012	.010	.001	.002	.001	.001	.002	.001
		.4	.016	.011	.002	.002	.001	.001	.002	.001
		.8	.021	.013	.004	.002	.001	.001	.002	.001
	60	.0	.001	.001	.000	.000	.000	.000	.000	.000
		.4	.001	.001	.000	.000	.000	.000	.000	.000
		.8	.002	.001	.000	.000	.000	.000	.000	.000
g-h-3	20	.0	.010	.010	.001	.001	.002	.002	.001	.001
		.4	.012	.010	.001	.001	.002	.001	.001	.001
		.8	.013	.010	.001	.001	.002	.001	.001	.001
	60	.0	.001	.001	.000	.000	.000	.000	.000	.000
		.4	.001	.001	.000	.000	.000	.000	.000	.000
		.8	.001	.001	.000	.000	.000	.000	.000	.000
g-h-4	20	.0	.014	.012	.003	.003	.001	.001	.003	.002
		.4	.018	.012	.004	.002	.000	.001	.003	.002
		.8	.024	.012	.006	.002	.000	.000	.005	.003
	60	.0	.001	.001	.000	.000	.000	.000	.000	.000
		.4	.002	.001	.000	.000	.000	.000	.000	.000
		.8	.002	.001	.000	.000	.000	.000	.000	.000
Hyper	20	.0	.003	.004	.001	.002	.000	.007	.008	.022
		.4	.005	.005	.001	.002	.000	.006	.007	.020
		.8	.004	.005	.001	.002	.000	.006	.006	.020
	60	.0	.000	.002	.001	.004	.000	.004	.002	.007
		.4	.000	.003	.001	.004	.001	.005	.003	.007
		.8	.000	.002	.001	.004	.000	.004	.002	.007

Anmerkung: Vorpalte 1 gibt die Verteilung an, Vorpalte 2 den Stichprobenumfang N , Vorpalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-14. MSE_{MC}

			$d_{s,R}^{S_1}$		$\hat{d}_R^{S_1}$		$d_{s,R}^{S_1,Boot}$		$\hat{d}_R^{S_1,Boot}$	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	.249	.256	.206	.211	.244	.245	.212	.213
		.4	.193	.221	.157	.181	.198	.217	.172	.190
		.8	.134	.175	.106	.142	.145	.178	.127	.155
	60	.0	.063	.063	.059	.059	.065	.064	.062	.061
		.4	.047	.054	.044	.050	.050	.056	.047	.053
		.8	.031	.044	.028	.041	.033	.046	.032	.044
Gleich	20	.0	.327	.342	.272	.285	.293	.290	.258	.254
		.4	.226	.281	.186	.233	.212	.244	.188	.215
		.8	.134	.210	.106	.172	.135	.186	.121	.165
	60	.0	.082	.080	.077	.076	.080	.078	.077	.075
		.4	.056	.067	.053	.064	.056	.066	.053	.063
		.8	.029	.050	.027	.047	.030	.050	.029	.048
g-h-1	20	.0	.185	.186	.151	.151	.199	.198	.173	.171
		.4	.156	.166	.126	.136	.176	.183	.152	.159
		.8	.128	.148	.102	.120	.150	.168	.130	.146
	60	.0	.046	.045	.043	.042	.050	.049	.047	.047
		.4	.038	.041	.035	.039	.042	.046	.040	.044
		.8	.030	.037	.028	.034	.034	.042	.033	.039
g-h-2	20	.0	.180	.134	.147	.109	.200	.134	.174	.116
		.4	.173	.121	.138	.097	.191	.127	.166	.110
		.8	.175	.107	.138	.084	.191	.114	.166	.099
	60	.0	.044	.032	.041	.030	.049	.035	.047	.033
		.4	.041	.027	.038	.026	.046	.030	.044	.029
		.8	.039	.024	.036	.022	.044	.027	.042	.026
g-h-3	20	.0	.162	.143	.132	.116	.181	.158	.157	.136
		.4	.146	.131	.117	.106	.167	.146	.144	.126
		.8	.133	.119	.105	.095	.156	.136	.134	.117
	60	.0	.039	.034	.037	.032	.044	.038	.042	.037
		.4	.034	.031	.032	.029	.039	.035	.037	.033
		.8	.030	.029	.028	.027	.035	.033	.033	.031
g-h-4	20	.0	.228	.189	.188	.156	.224	.164	.196	.142
		.4	.202	.159	.163	.130	.209	.144	.184	.126
		.8	.195	.120	.154	.096	.202	.116	.179	.103
	60	.0	.053	.044	.049	.041	.055	.044	.052	.042
		.4	.045	.036	.042	.033	.049	.037	.046	.035
		.8	.038	.027	.035	.025	.041	.029	.039	.028
Hyper	20	.0	.285	.609	.245	.525	.349	.701	.310	.622
		.4	.240	.527	.204	.454	.316	.633	.280	.563
		.8	.183	.394	.156	.339	.264	.450	.234	.404
	60	.0	.077	.096	.074	.093	.098	.094	.096	.093
		.4	.059	.082	.057	.081	.079	.082	.078	.082
		.8	.044	.065	.043	.064	.066	.067	.065	.066

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-15. \overline{B}_{MC}^2

			$d_{s,R}^{S_2}$		$\hat{d}_R^{S_2}$		$d_{s,R}^{S_2,Boot}$		$\hat{d}_R^{S_2,Boot}$	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	.012	.012	.001	.001	.001	.001	.002	.002
		.4	.011	.011	.001	.001	.000	.000	.003	.003
		.8	.012	.013	.001	.001	.001	.001	.002	.002
	60	.0	.001	.001	.000	.000	.000	.000	.000	.000
		.4	.001	.001	.000	.000	.000	.000	.000	.000
		.8	.001	.001	.000	.000	.000	.000	.000	.000
Gleich	20	.0	.014	.015	.002	.002	.000	.000	.004	.004
		.4	.013	.013	.001	.002	.000	.000	.005	.004
		.8	.013	.013	.001	.002	.000	.000	.004	.004
	60	.0	.001	.002	.000	.000	.000	.000	.000	.000
		.4	.001	.001	.000	.000	.000	.000	.000	.000
		.8	.001	.001	.000	.000	.000	.000	.000	.000
g-h-1	20	.0	.011	.011	.001	.001	.001	.001	.001	.002
		.4	.012	.011	.001	.001	.002	.001	.001	.001
		.8	.012	.011	.001	.001	.001	.001	.001	.001
	60	.0	.001	.001	.000	.000	.000	.000	.000	.000
		.4	.001	.001	.000	.000	.000	.000	.000	.000
		.8	.001	.001	.000	.000	.000	.000	.000	.000
g-h-2	20	.0	.037	.025	.012	.008	.000	.000	.003	.003
		.4	.028	.020	.008	.006	.000	.000	.004	.002
		.8	.027	.016	.007	.004	.001	.000	.002	.001
	60	.0	.003	.002	.001	.001	.000	.000	.000	.000
		.4	.003	.002	.001	.000	.000	.000	.000	.000
		.8	.002	.001	.001	.000	.000	.000	.000	.000
g-h-3	20	.0	.014	.012	.002	.002	.001	.001	.001	.001
		.4	.014	.011	.002	.001	.001	.001	.001	.001
		.8	.014	.010	.002	.001	.002	.001	.001	.001
	60	.0	.002	.001	.000	.000	.001	.000	.000	.000
		.4	.002	.001	.000	.000	.001	.000	.000	.000
		.8	.001	.001	.000	.000	.000	.000	.000	.000
g-h-4	20	.0	.049	.037	.019	.015	.001	.000	.009	.005
		.4	.040	.024	.014	.007	.000	.000	.007	.005
		.8	.033	.017	.010	.004	.000	.000	.006	.003
	60	.0	.005	.003	.002	.001	.000	.000	.000	.000
		.4	.003	.002	.001	.001	.000	.000	.000	.000
		.8	.002	.002	.001	.000	.000	.000	.000	.000
Hyper	20	.0	.018	.040	.002	.014	.026	.024	.011	.007
		.4	.018	.041	.002	.015	.025	.027	.010	.008
		.8	.015	.037	.001	.014	.022	.026	.008	.008
	60	.0	.007	.011	.004	.007	.004	.007	.002	.006
		.4	.005	.010	.003	.006	.003	.006	.002	.005
		.8	.005	.009	.003	.006	.004	.006	.003	.005

Anmerkung: Vorpalte 1 gibt die Verteilung an, Vorpalte 2 den Stichprobenumfang N , Vorpalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-16. MSE_{MC}

			$d_{s,R}^{S_2}$		$\hat{d}_R^{S_2}$		$d_{s,R}^{S_2,Boot}$		$\hat{d}_R^{S_2,Boot}$	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	.249	.254	.206	.210	.247	.254	.215	.221
		.4	.191	.209	.156	.172	.192	.212	.169	.185
		.8	.131	.183	.104	.148	.141	.190	.123	.166
	60	.0	.064	.062	.061	.059	.066	.065	.063	.062
		.4	.047	.054	.044	.051	.050	.057	.047	.054
		.8	.031	.043	.028	.040	.033	.045	.032	.043
Gleich	20	.0	.330	.320	.275	.266	.296	.301	.260	.264
		.4	.229	.255	.189	.211	.213	.245	.189	.216
		.8	.134	.208	.106	.170	.133	.198	.120	.175
	60	.0	.083	.080	.079	.075	.081	.079	.078	.076
		.4	.056	.066	.053	.063	.056	.067	.053	.064
		.8	.030	.049	.027	.046	.030	.049	.029	.047
g-h-1	20	.0	.186	.195	.153	.160	.203	.210	.176	.182
		.4	.163	.167	.132	.136	.182	.184	.157	.159
		.8	.128	.156	.102	.126	.149	.176	.129	.153
	60	.0	.047	.046	.044	.043	.052	.050	.049	.048
		.4	.038	.041	.036	.038	.043	.045	.041	.043
		.8	.030	.036	.027	.034	.034	.041	.032	.039
g-h-2	20	.0	.432	.295	.354	.242	.407	.273	.355	.239
		.4	.315	.208	.256	.168	.315	.207	.276	.181
		.8	.230	.142	.183	.113	.248	.155	.216	.135
	60	.0	.095	.065	.089	.061	.096	.065	.092	.062
		.4	.071	.049	.066	.046	.074	.051	.071	.049
		.8	.051	.032	.047	.030	.055	.036	.052	.034
g-h-3	20	.0	.210	.188	.171	.153	.226	.199	.196	.173
		.4	.174	.155	.140	.126	.191	.172	.165	.149
		.8	.144	.132	.115	.106	.170	.152	.146	.132
	60	.0	.052	.045	.049	.042	.057	.049	.054	.047
		.4	.042	.038	.039	.036	.047	.042	.044	.040
		.8	.032	.030	.030	.028	.038	.034	.036	.032
g-h-4	20	.0	.622	.436	.515	.361	.534	.384	.470	.337
		.4	.461	.284	.378	.233	.424	.266	.374	.235
		.8	.284	.172	.227	.139	.287	.179	.254	.158
	60	.0	.124	.090	.117	.084	.118	.086	.113	.082
		.4	.087	.063	.081	.059	.085	.063	.082	.061
		.8	.054	.037	.050	.034	.056	.039	.054	.037
Hyper	20	.0	.614	.516	.517	.426	.995	.713	.849	.602
		.4	.505	.446	.424	.365	.873	.636	.743	.535
		.8	.379	.370	.316	.301	.710	.542	.603	.455
	60	.0	.300	.167	.284	.157	.492	.228	.469	.217
		.4	.246	.137	.233	.129	.425	.195	.406	.185
		.8	.205	.108	.193	.101	.373	.157	.356	.149

Anmerkung: Vorderspalte 1 gibt die Verteilung an, Vorderspalte 2 den Stichprobenumfang N , Vorderspalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-17. $|\overline{CP-.95}|$

			Perc ($d_{D,s}$)		BCA ($d_{D,s}$)		Perc-t ($d_{D,s}$)		Perc (\hat{d}_D)		BCA (\hat{d}_D)		Perc-t (\hat{d}_D)		NC ($d_{D,s}$)	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	.042	.042	.011	.016	.007	.009	.035	.035	.012	.016	.012	.013	.003	.002
		.4	.041	.040	.016	.015	.009	.008	.033	.033	.018	.017	.015	.013	.003	.003
		.8	.054	.047	.026	.020	.019	.010	.039	.037	.028	.019	.031	.016	.002	.001
	60	.0	.016	.013	.005	.005	.005	.004	.014	.012	.006	.005	.006	.005	.003	.002
		.4	.016	.016	.006	.007	.005	.007	.013	.014	.006	.008	.006	.009	.002	.003
		.8	.021	.016	.011	.006	.008	.005	.017	.012	.012	.007	.011	.008	.002	.001
Gleich	20	.0	.033	.026	.009	.009	.005	.013	.029	.023	.011	.009	.008	.014	.005	.009
		.4	.043	.022	.015	.007	.011	.014	.036	.019	.018	.010	.017	.018	.002	.009
		.8	.074	.019	.034	.010	.033	.020	.060	.016	.036	.013	.043	.029	.015	.016
	60	.0	.011	.008	.002	.002	.003	.006	.009	.008	.003	.004	.004	.007	.005	.007
		.4	.014	.006	.005	.002	.007	.005	.012	.006	.007	.004	.008	.006	.002	.012
		.8	.025	.004	.009	.003	.009	.008	.020	.005	.010	.005	.013	.011	.012	.017
g-h-1	20	.0	.096	.117	.042	.052	.055	.066	.081	.100	.037	.046	.055	.065	.035	.046
		.4	.125	.138	.060	.070	.065	.084	.106	.119	.051	.061	.063	.082	.051	.067
		.8	.192	.177	.112	.095	.105	.106	.162	.155	.094	.084	.097	.103	.108	.097
	60	.0	.056	.072	.030	.040	.048	.060	.051	.065	.027	.037	.048	.059	.055	.069
		.4	.081	.093	.048	.053	.066	.081	.072	.085	.044	.049	.064	.078	.083	.101
		.8	.147	.117	.089	.067	.117	.096	.133	.106	.081	.063	.112	.091	.174	.137
g-h-2	20	.0	.080	.058	.031	.025	.045	.027	.070	.043	.028	.025	.047	.032	.025	.028
		.4	.100	.074	.045	.036	.055	.042	.086	.053	.040	.033	.058	.048	.041	.031
		.8	.136	.117	.068	.066	.074	.064	.114	.085	.062	.056	.079	.074	.087	.031
	60	.0	.036	.023	.016	.015	.025	.019	.032	.018	.015	.016	.026	.022	.030	.029
		.4	.048	.045	.023	.024	.032	.034	.042	.036	.022	.022	.033	.036	.049	.029
		.8	.061	.067	.031	.040	.038	.054	.053	.056	.029	.036	.039	.057	.093	.029
g-h-3	20	.0	.123	.120	.058	.056	.073	.061	.107	.102	.050	.048	.069	.059	.050	.029
		.4	.168	.157	.090	.078	.095	.077	.145	.133	.079	.066	.089	.073	.083	.046
		.8	.267	.202	.164	.111	.157	.110	.236	.175	.142	.096	.144	.103	.171	.075
	60	.0	.083	.083	.047	.046	.071	.066	.076	.074	.044	.042	.069	.064	.085	.050
		.4	.128	.127	.078	.075	.106	.099	.118	.116	.071	.067	.101	.094	.140	.089
		.8	.223	.160	.148	.097	.188	.124	.209	.148	.138	.090	.180	.120	.276	.131
g-h-4	20	.0	.042	.025	.013	.012	.015	.015	.036	.019	.014	.014	.018	.018	.003	.029
		.4	.060	.026	.020	.009	.025	.016	.049	.016	.020	.012	.031	.027	.012	.031
		.8	.088	.022	.039	.013	.038	.028	.071	.012	.039	.018	.049	.046	.038	.035
	60	.0	.012	.008	.002	.003	.004	.008	.011	.007	.004	.005	.005	.009	.002	.029
		.4	.019	.009	.007	.004	.008	.009	.016	.007	.007	.005	.010	.012	.010	.034
		.8	.034	.005	.014	.004	.017	.012	.029	.004	.016	.007	.020	.017	.037	.036
Hyper	20	.0	.048	.019	.020	.009	.020	.016	.043	.018	.019	.009	.023	.016	.022	.010
		.4	.053	.020	.021	.006	.014	.016	.043	.018	.021	.008	.019	.018	.010	.012
		.8	.055	.015	.024	.009	.023	.020	.042	.014	.027	.012	.034	.028	.009	.014
	60	.0	.016	.007	.008	.002	.011	.010	.014	.006	.009	.002	.012	.008	.022	.011
		.4	.018	.007	.008	.003	.008	.011	.014	.007	.008	.005	.008	.013	.011	.013
		.8	.016	.007	.006	.002	.007	.009	.013	.007	.008	.004	.010	.014	.009	.017

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-18. Mittlere Intervalllänge

			Perc ($d_{D,s}$)		BCA ($d_{D,s}$)		Perc-t ($d_{D,s}$)		Perc (\hat{d}_D)		BCA (\hat{d}_D)		Perc-t (\hat{d}_D)		NC ($d_{D,s}$)	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	1.522	1.525	1.461	1.464	1.385	1.387	1.461	1.464	1.403	1.405	1.330	1.331	1.407	1.407
		.4	1.251	1.368	1.185	1.304	1.127	1.238	1.201	1.314	1.138	1.252	1.081	1.189	1.154	1.262
		.8	.896	1.190	.818	1.121	.783	1.067	.860	1.142	.785	1.076	.752	1.024	.823	1.098
	60	.0	.818	.816	.809	.808	.797	.796	.807	.806	.799	.798	.787	.786	.799	.799
		.4	.669	.732	.659	.722	.650	.712	.661	.722	.651	.713	.642	.703	.655	.716
		.8	.474	.636	.461	.626	.456	.617	.468	.628	.455	.618	.450	.609	.464	.622
Gleich	20	.0	1.505	1.492	1.449	1.437	1.377	1.368	1.445	1.432	1.391	1.380	1.322	1.313	1.403	1.401
		.4	1.256	1.320	1.188	1.264	1.130	1.206	1.206	1.267	1.141	1.213	1.085	1.158	1.154	1.256
		.8	.964	1.115	.872	1.059	.824	1.015	.925	1.070	.837	1.017	.791	.975	.828	1.089
	60	.0	.801	.787	.793	.780	.782	.769	.791	.777	.783	.770	.772	.759	.798	.798
		.4	.667	.696	.657	.689	.648	.680	.658	.687	.648	.680	.640	.671	.655	.715
		.8	.505	.585	.489	.577	.483	.570	.499	.577	.482	.570	.477	.563	.465	.621
g-h-1	20	.0	1.642	1.687	1.554	1.584	1.445	1.466	1.576	1.619	1.491	1.521	1.388	1.407	1.436	1.445
		.4	1.408	1.581	1.312	1.470	1.211	1.347	1.351	1.518	1.259	1.411	1.163	1.293	1.197	1.311
		.8	1.130	1.477	1.026	1.349	.915	1.216	1.085	1.418	.985	1.295	.879	1.168	.883	1.162
	60	.0	.914	.937	.902	.922	.875	.891	.903	.925	.890	.910	.864	.880	.807	.810
		.4	.789	.889	.776	.873	.746	.835	.779	.878	.766	.862	.736	.825	.667	.730
		.8	.655	.838	.651	.818	.598	.774	.647	.828	.643	.808	.591	.764	.483	.640
g-h-2	20	.0	1.637	1.298	1.547	1.247	1.441	1.185	1.571	1.246	1.485	1.197	1.383	1.138	1.422	1.408
		.4	1.450	1.159	1.334	1.106	1.228	1.035	1.392	1.112	1.281	1.062	1.179	.993	1.178	1.269
		.8	1.216	1.020	1.070	.973	.950	.878	1.167	.979	1.027	.934	.912	.843	.858	1.115
	60	.0	.893	.677	.880	.668	.859	.659	.882	.668	.868	.659	.849	.651	.802	.800
		.4	.780	.597	.761	.591	.741	.577	.770	.590	.752	.584	.732	.570	.661	.718
		.8	.641	.528	.615	.529	.591	.503	.633	.521	.607	.522	.583	.497	.470	.626
g-h-3	20	.0	1.695	1.580	1.594	1.493	1.473	1.387	1.627	1.517	1.530	1.433	1.414	1.331	1.450	1.446
		.4	1.485	1.475	1.373	1.380	1.252	1.264	1.426	1.416	1.318	1.325	1.202	1.214	1.219	1.318
		.8	1.248	1.381	1.125	1.273	.981	1.139	1.198	1.325	1.080	1.222	.942	1.093	.925	1.171
	60	.0	.952	.863	.936	.856	.903	.822	.940	.852	.924	.845	.891	.811	.812	.812
		.4	.850	.814	.835	.808	.790	.764	.839	.803	.824	.797	.780	.754	.676	.735
		.8	.742	.773	.739	.765	.660	.709	.733	.763	.730	.756	.652	.700	.499	.646
g-h-4	20	.0	1.545	1.258	1.482	1.212	1.400	1.173	1.483	1.207	1.423	1.164	1.344	1.126	1.407	1.395
		.4	1.335	1.092	1.248	1.040	1.174	1.009	1.282	1.048	1.198	.998	1.127	.969	1.160	1.251
		.8	1.083	.904	.959	.845	.887	.822	1.039	.868	.921	.812	.851	.789	.837	1.086
	60	.0	.823	.657	.814	.652	.802	.646	.813	.649	.804	.643	.792	.638	.799	.797
		.4	.704	.561	.690	.554	.680	.551	.695	.554	.681	.547	.671	.544	.656	.714
		.8	.553	.450	.530	.441	.523	.441	.546	.444	.524	.436	.516	.435	.466	.620
Hyper	20	.0	1.620	1.460	1.560	1.414	1.463	1.346	1.555	1.402	1.497	1.357	1.404	1.292	1.408	1.399
		.4	1.281	1.282	1.223	1.234	1.153	1.178	1.230	1.231	1.174	1.184	1.107	1.131	1.156	1.254
		.8	.643	1.102	.599	1.051	.572	1.004	.617	1.058	.575	1.009	.549	.964	.600	1.088
	60	.0	.892	.785	.880	.779	.865	.768	.881	.775	.869	.769	.854	.758	.801	.798
		.4	.700	.685	.690	.679	.679	.670	.691	.677	.681	.670	.670	.661	.656	.715
		.8	.485	.584	.471	.577	.466	.569	.479	.576	.465	.569	.460	.562	.466	.620

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-19. $|\overline{CP} - .95|$

			Perc (d'_s)		BCA (d'_s)		Perc-t ₁ (d'_s)		Perc-t ₂ (d'_s)	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	.034	.039	.011	.012	.015	.011	.008	.005
		.4	.028	.037	.009	.013	.010	.011	.006	.006
		.8	.038	.044	.016	.016	.017	.011	.015	.007
	60	.0	.013	.012	.004	.005	.006	.005	.004	.004
		.4	.009	.015	.003	.007	.003	.007	.003	.006
		.8	.012	.015	.007	.006	.006	.004	.006	.004
Gleich	20	.0	.024	.022	.008	.007	.004	.011	.005	.017
		.4	.022	.018	.009	.006	.006	.013	.003	.018
		.8	.017	.018	.011	.009	.014	.019	.013	.020
	60	.0	.008	.007	.002	.003	.001	.005	.003	.006
		.4	.006	.006	.003	.001	.002	.005	.001	.006
		.8	.004	.004	.004	.004	.003	.007	.002	.007
g-h-1	20	.0	.088	.117	.041	.053	.071	.080	.062	.068
		.4	.110	.124	.053	.061	.080	.088	.073	.078
		.8	.164	.156	.089	.083	.110	.097	.101	.090
	60	.0	.051	.072	.030	.040	.055	.068	.049	.062
		.4	.069	.080	.040	.047	.067	.077	.063	.071
		.8	.107	.098	.063	.057	.093	.085	.089	.083
g-h-2	20	.0	.073	.050	.035	.023	.057	.027	.053	.025
		.4	.086	.059	.046	.029	.066	.038	.064	.036
		.8	.129	.096	.075	.056	.090	.056	.087	.053
	60	.0	.034	.021	.017	.014	.029	.022	.027	.021
		.4	.040	.037	.022	.023	.035	.030	.034	.030
		.8	.060	.055	.038	.036	.049	.048	.049	.046
g-h-3	20	.0	.116	.117	.058	.055	.092	.070	.083	.060
		.4	.145	.137	.075	.070	.105	.082	.095	.073
		.8	.215	.173	.125	.096	.146	.097	.135	.090
	60	.0	.080	.083	.047	.045	.082	.072	.076	.067
		.4	.103	.104	.061	.060	.099	.086	.093	.082
		.8	.159	.131	.098	.077	.137	.105	.131	.103
g-h-4	20	.0	.033	.018	.012	.007	.020	.016	.020	.019
		.4	.035	.018	.017	.006	.025	.015	.024	.019
		.8	.046	.018	.027	.011	.036	.029	.036	.030
	60	.0	.013	.006	.002	.002	.004	.008	.003	.009
		.4	.009	.007	.002	.002	.007	.008	.006	.008
		.8	.012	.003	.003	.003	.012	.012	.011	.012
Hyper	20	.0	.035	.016	.017	.009	.022	.017	.017	.020
		.4	.036	.016	.017	.010	.019	.016	.017	.016
		.8	.041	.014	.023	.010	.022	.017	.021	.018
	60	.0	.011	.005	.005	.002	.008	.010	.009	.011
		.4	.010	.006	.005	.002	.006	.011	.007	.012
		.8	.012	.008	.005	.003	.005	.009	.006	.009

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-20. Mittlere Intervalllänge

			Perc (d'_s)		BCA (d'_s)		Perc- t_1 (d'_s)		Perc- t_2 (d'_s)	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	1.369	1.457	1.349	1.418	1.319	1.359	1.335	1.381
		.4	1.125	1.309	1.096	1.263	1.076	1.212	1.082	1.229
		.8	.833	1.161	.778	1.102	.762	1.054	.766	1.065
	60	.0	.768	.798	.765	.792	.760	.782	.764	.785
		.4	.624	.715	.620	.708	.617	.699	.618	.702
		.8	.449	.628	.440	.619	.438	.611	.439	.613
Gleich	20	.0	1.347	1.433	1.329	1.395	1.303	1.342	1.319	1.360
		.4	1.068	1.269	1.048	1.229	1.034	1.183	1.038	1.198
		.8	.726	1.096	.692	1.048	.683	1.009	.685	1.016
	60	.0	.748	.768	.746	.763	.741	.755	.745	.758
		.4	.591	.679	.589	.674	.586	.666	.588	.668
		.8	.387	.580	.382	.574	.380	.567	.381	.568
g-h-1	20	.0	1.483	1.592	1.445	1.527	1.379	1.428	1.404	1.463
		.4	1.287	1.466	1.230	1.388	1.163	1.291	1.178	1.320
		.8	1.107	1.378	1.012	1.279	.915	1.168	.928	1.185
	60	.0	.868	.914	.864	.906	.844	.876	.852	.885
		.4	.750	.845	.743	.836	.720	.805	.725	.813
		.8	.643	.790	.635	.777	.594	.742	.596	.745
g-h-2	20	.0	1.478	1.227	1.436	1.194	1.372	1.153	1.386	1.170
		.4	1.311	1.094	1.246	1.057	1.175	1.017	1.184	1.034
		.8	1.146	.985	1.037	.943	.922	.879	.932	.891
	60	.0	.845	.653	.841	.647	.826	.640	.830	.644
		.4	.735	.579	.725	.574	.709	.567	.711	.570
		.8	.624	.520	.603	.517	.578	.501	.580	.503
g-h-3	20	.0	1.530	1.495	1.485	1.441	1.404	1.353	1.432	1.382
		.4	1.338	1.364	1.272	1.298	1.188	1.212	1.206	1.238
		.8	1.185	1.283	1.076	1.197	.953	1.093	.967	1.107
	60	.0	.903	.841	.900	.839	.871	.807	.882	.814
		.4	.794	.771	.788	.770	.753	.736	.761	.743
		.8	.699	.726	.691	.720	.633	.680	.637	.683
g-h-4	20	.0	1.387	1.196	1.363	1.164	1.328	1.140	1.333	1.154
		.4	1.166	1.046	1.124	1.006	1.097	.992	1.096	1.004
		.8	.921	.890	.836	.840	.802	.826	.806	.834
	60	.0	.772	.634	.768	.630	.763	.627	.766	.628
		.4	.633	.547	.627	.542	.623	.541	.624	.542
		.8	.465	.453	.451	.446	.448	.446	.449	.447
Hyper	20	.0	1.501	1.413	1.473	1.380	1.430	1.326	1.442	1.341
		.4	1.248	1.246	1.213	1.212	1.188	1.168	1.187	1.180
		.8	.975	1.075	.913	1.037	.891	.996	.886	1.002
	60	.0	.853	.768	.846	.764	.839	.755	.842	.757
		.4	.702	.675	.693	.670	.690	.663	.690	.665
		.8	.539	.577	.525	.572	.523	.565	.522	.566

Anmerkung: Vorderspalte 1 gibt die Verteilung an, Vorderspalte 2 den Stichprobenumfang N , Vorderspalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-21. $|\overline{CP} - .95|$

			Perc (\hat{d}'_{cl})		BCA (\hat{d}'_{cl})		Perc-t (\hat{d}'_{cl})		NC ₁ (d'_s)	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	.030	.033	.013	.013	.017	.013	.008	.005
		.4	.023	.030	.011	.014	.011	.015	.012	.003
		.8	.025	.033	.021	.019	.027	.017	.007	.003
	60	.0	.011	.010	.004	.005	.006	.007	.009	.005
		.4	.007	.013	.003	.007	.005	.009	.012	.003
		.8	.008	.012	.009	.007	.009	.006	.008	.002
Gleich	20	.0	.022	.020	.011	.008	.006	.012	.010	.012
		.4	.020	.018	.014	.008	.013	.015	.015	.012
		.8	.014	.015	.022	.013	.034	.028	.024	.016
	60	.0	.008	.008	.003	.003	.001	.006	.013	.010
		.4	.006	.006	.004	.002	.003	.006	.017	.014
		.8	.004	.003	.005	.002	.009	.009	.026	.017
g-h-1	20	.0	.074	.099	.036	.046	.068	.077	.023	.039
		.4	.090	.105	.045	.053	.077	.085	.036	.050
		.8	.138	.135	.075	.072	.104	.093	.096	.073
	60	.0	.046	.065	.028	.036	.052	.065	.042	.067
		.4	.060	.072	.036	.044	.064	.074	.067	.081
		.8	.095	.089	.058	.053	.088	.082	.143	.110
g-h-2	20	.0	.063	.037	.033	.021	.054	.031	.015	.031
		.4	.072	.042	.042	.026	.066	.043	.027	.032
		.8	.108	.071	.068	.048	.093	.067	.079	.031
	60	.0	.029	.016	.016	.013	.028	.024	.018	.031
		.4	.036	.030	.021	.022	.034	.032	.036	.031
		.8	.053	.044	.037	.033	.050	.049	.092	.028
g-h-3	20	.0	.099	.100	.051	.048	.087	.066	.036	.023
		.4	.123	.114	.064	.059	.097	.077	.058	.032
		.8	.185	.147	.108	.083	.138	.092	.138	.055
	60	.0	.072	.074	.043	.041	.079	.069	.073	.046
		.4	.093	.093	.056	.054	.093	.083	.108	.067
		.8	.146	.119	.091	.071	.130	.101	.212	.101
g-h-4	20	.0	.029	.015	.014	.011	.022	.020	.006	.030
		.4	.030	.014	.019	.009	.029	.025	.006	.031
		.8	.037	.013	.032	.017	.049	.046	.003	.034
	60	.0	.011	.006	.003	.004	.005	.009	.008	.030
		.4	.007	.007	.002	.005	.007	.011	.010	.033
		.8	.009	.003	.006	.007	.016	.018	.008	.036
Hyper	20	.0	.033	.015	.018	.009	.023	.016	.014	.012
		.4	.031	.016	.020	.010	.023	.017	.015	.012
		.8	.035	.014	.028	.012	.029	.025	.031	.015
	60	.0	.009	.005	.005	.001	.009	.008	.015	.013
		.4	.008	.006	.006	.003	.007	.011	.014	.015
		.8	.010	.007	.006	.004	.008	.011	.025	.018

Anmerkung: Vorpalte 1 gibt die Verteilung an, Vorpalte 2 den Stichprobenumfang N , Vorpalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-22. Mittlere Intervalllänge

			Perc (\hat{d}'_{cl})		BCA (\hat{d}'_{cl})		Perc-t (\hat{d}'_{cl})		NC ₁ (d'_s)	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	1.314	1.399	1.295	1.361	1.266	1.304	1.391	1.402
		.4	1.080	1.257	1.052	1.213	1.033	1.163	1.151	1.257
		.8	.800	1.115	.746	1.058	.732	1.012	.826	1.097
	60	.0	.758	.787	.755	.782	.750	.772	.796	.798
		.4	.616	.706	.612	.699	.609	.690	.655	.716
		.8	.444	.620	.435	.611	.432	.604	.464	.622
Gleich	20	.0	1.293	1.376	1.276	1.339	1.251	1.288	1.387	1.396
		.4	1.025	1.218	1.006	1.180	.992	1.136	1.140	1.250
		.8	.697	1.052	.665	1.006	.656	.969	.812	1.089
	60	.0	.738	.758	.736	.754	.732	.745	.796	.798
		.4	.584	.670	.581	.665	.579	.658	.653	.714
		.8	.382	.573	.377	.566	.375	.560	.462	.621
g-h-1	20	.0	1.424	1.528	1.387	1.466	1.323	1.371	1.420	1.440
		.4	1.235	1.407	1.181	1.332	1.116	1.239	1.181	1.284
		.8	1.063	1.323	.971	1.228	.879	1.122	.875	1.135
	60	.0	.857	.902	.853	.894	.833	.865	.804	.809
		.4	.740	.834	.734	.825	.711	.795	.661	.723
		.8	.635	.780	.627	.767	.586	.732	.477	.631
g-h-2	20	.0	1.419	1.178	1.379	1.147	1.317	1.106	1.406	1.403
		.4	1.259	1.050	1.196	1.015	1.128	.977	1.175	1.257
		.8	1.100	.945	.995	.905	.885	.844	.867	1.106
	60	.0	.834	.645	.830	.639	.815	.632	.800	.799
		.4	.726	.572	.715	.567	.700	.560	.659	.716
		.8	.616	.513	.596	.511	.570	.495	.473	.624
g-h-3	20	.0	1.469	1.435	1.425	1.384	1.347	1.299	1.436	1.441
		.4	1.284	1.309	1.222	1.247	1.140	1.164	1.191	1.284
		.8	1.138	1.231	1.033	1.149	.915	1.049	.894	1.137
	60	.0	.892	.830	.888	.828	.860	.796	.810	.811
		.4	.784	.761	.778	.760	.744	.727	.664	.723
		.8	.690	.717	.682	.711	.625	.671	.483	.633
g-h-4	20	.0	1.332	1.148	1.308	1.118	1.275	1.095	1.391	1.390
		.4	1.119	1.004	1.079	.966	1.053	.952	1.148	1.247
		.8	.885	.854	.802	.806	.769	.793	.828	1.086
	60	.0	.762	.626	.759	.622	.753	.619	.797	.796
		.4	.625	.540	.619	.535	.615	.534	.653	.714
		.8	.459	.447	.445	.440	.443	.440	.463	.620
Hyper	20	.0	1.441	1.357	1.414	1.325	1.373	1.273	1.392	1.394
		.4	1.198	1.196	1.164	1.164	1.141	1.121	1.151	1.249
		.8	.936	1.032	.877	.995	.856	.956	.829	1.088
	60	.0	.842	.758	.835	.754	.828	.745	.798	.797
		.4	.693	.666	.685	.662	.681	.655	.655	.714
		.8	.532	.570	.518	.564	.516	.558	.467	.620

Anmerkung: Vorpalte 1 gibt die Verteilung an, Vorpalte 2 den Stichprobenumfang N , Vorpalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-23. $|\overline{CP} - .95|$

			Perc (\hat{d}'_{c_2})		BCA (\hat{d}'_{c_2})		Perc-t (\hat{d}'_{c_2})		NC ₂ (d'_s)	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	.031	.034	.013	.013	.010	.008	.008	.002
		.4	.024	.031	.010	.014	.008	.010	.004	.003
		.8	.027	.034	.019	.018	.024	.013	.004	.002
	60	.0	.011	.010	.004	.005	.004	.005	.004	.001
		.4	.008	.013	.003	.007	.004	.007	.004	.003
		.8	.009	.012	.009	.007	.009	.005	.006	.001
Gleich	20	.0	.023	.020	.009	.007	.005	.017	.007	.009
		.4	.020	.018	.011	.007	.008	.018	.008	.008
		.8	.013	.015	.018	.013	.029	.029	.020	.014
	60	.0	.008	.008	.002	.003	.003	.007	.006	.008
		.4	.006	.006	.003	.002	.002	.007	.009	.011
		.8	.004	.003	.005	.002	.007	.009	.023	.015
g-h-1	20	.0	.079	.102	.038	.048	.061	.067	.044	.054
		.4	.097	.107	.049	.054	.072	.077	.066	.064
		.8	.143	.136	.077	.073	.096	.086	.122	.080
	60	.0	.047	.066	.028	.037	.049	.060	.064	.076
		.4	.063	.074	.038	.044	.062	.069	.091	.092
		.8	.097	.090	.059	.053	.085	.080	.166	.117
g-h-2	20	.0	.067	.039	.034	.021	.053	.031	.032	.029
		.4	.077	.044	.043	.026	.066	.041	.050	.030
		.8	.111	.073	.069	.048	.090	.063	.103	.029
	60	.0	.031	.017	.016	.013	.027	.024	.034	.030
		.4	.038	.031	.022	.022	.034	.032	.057	.029
		.8	.055	.045	.037	.033	.050	.048	.114	.029
g-h-3	20	.0	.105	.103	.056	.049	.081	.059	.065	.036
		.4	.129	.118	.068	.061	.092	.070	.091	.044
		.8	.189	.148	.111	.084	.130	.086	.168	.061
	60	.0	.075	.076	.045	.042	.076	.066	.098	.055
		.4	.097	.095	.058	.055	.090	.080	.140	.077
		.8	.149	.120	.092	.071	.126	.099	.238	.105
g-h-4	20	.0	.031	.015	.013	.009	.021	.023	.005	.030
		.4	.031	.014	.018	.008	.027	.026	.008	.031
		.8	.039	.013	.031	.017	.047	.046	.009	.034
	60	.0	.011	.006	.002	.003	.004	.011	.003	.030
		.4	.008	.006	.002	.004	.006	.011	.003	.033
		.8	.009	.004	.005	.006	.015	.018	.005	.036
Hyper	20	.0	.033	.015	.018	.009	.018	.018	.031	.010
		.4	.031	.016	.019	.010	.018	.017	.032	.011
		.8	.036	.014	.027	.012	.027	.024	.044	.014
	60	.0	.009	.005	.005	.002	.010	.009	.029	.011
		.4	.008	.006	.005	.003	.008	.011	.032	.011
		.8	.011	.007	.005	.004	.007	.011	.041	.017

Anmerkung: Vorpalte 1 gibt die Verteilung an, Vorpalte 2 den Stichprobenumfang N , Vorpalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-24. Mittlere Intervalllänge

			Perc (\hat{d}'_{c_2})		BCA (\hat{d}'_{c_2})		Perc-t (\hat{d}'_{c_2})		NC ₂ (d'_s)	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	1.338	1.413	1.318	1.374	1.307	1.338	1.322	1.374
		.4	1.099	1.269	1.071	1.224	1.058	1.190	1.082	1.231
		.8	.812	1.121	.757	1.063	.746	1.027	.780	1.084
	60	.0	.763	.790	.760	.784	.759	.778	.759	.784
		.4	.620	.707	.616	.701	.614	.695	.615	.703
		.8	.446	.621	.437	.612	.435	.606	.439	.616
Gleich	20	.0	1.317	1.389	1.300	1.351	1.292	1.317	1.318	1.370
		.4	1.043	1.228	1.025	1.190	1.016	1.159	1.071	1.227
		.8	.707	1.057	.675	1.011	.667	.980	.768	1.077
	60	.0	.743	.760	.741	.756	.740	.750	.758	.784
		.4	.587	.672	.585	.667	.584	.661	.614	.702
		.8	.384	.573	.379	.567	.378	.562	.436	.614
g-h-1	20	.0	1.448	1.546	1.411	1.481	1.374	1.420	1.346	1.401
		.4	1.257	1.422	1.200	1.345	1.151	1.279	1.106	1.250
		.8	1.077	1.331	.982	1.234	.902	1.142	.827	1.119
	60	.0	.862	.906	.859	.897	.846	.877	.765	.792
		.4	.744	.837	.738	.828	.721	.805	.620	.707
		.8	.637	.781	.629	.768	.591	.737	.451	.624
g-h-2	20	.0	1.440	1.188	1.399	1.156	1.355	1.132	1.339	1.374
		.4	1.278	1.059	1.214	1.023	1.157	1.000	1.107	1.230
		.8	1.114	.951	1.006	.910	.907	.860	.819	1.092
	60	.0	.839	.646	.835	.640	.824	.636	.762	.785
		.4	.730	.573	.720	.568	.706	.564	.620	.702
		.8	.619	.514	.598	.511	.575	.498	.447	.617
g-h-3	20	.0	1.494	1.450	1.449	1.398	1.400	1.340	1.359	1.404
		.4	1.306	1.323	1.240	1.257	1.179	1.199	1.114	1.250
		.8	1.152	1.238	1.044	1.154	.940	1.068	.844	1.121
	60	.0	.897	.833	.894	.831	.876	.806	.769	.794
		.4	.789	.763	.783	.762	.756	.736	.622	.707
		.8	.693	.718	.684	.712	.631	.675	.457	.626
g-h-4	20	.0	1.353	1.157	1.330	1.125	1.305	1.115	1.324	1.365
		.4	1.137	1.011	1.097	.973	1.072	.970	1.081	1.224
		.8	.897	.859	.813	.810	.784	.804	.782	1.074
	60	.0	.766	.627	.763	.623	.761	.621	.759	.783
		.4	.629	.541	.623	.536	.620	.536	.614	.701
		.8	.461	.448	.447	.440	.445	.441	.438	.614
Hyper	20	.0	1.465	1.368	1.439	1.336	1.411	1.297	1.326	1.369
		.4	1.218	1.206	1.184	1.173	1.162	1.141	1.084	1.227
		.8	.947	1.036	.887	1.000	.862	.965	.784	1.077
	60	.0	.847	.760	.840	.756	.836	.749	.760	.784
		.4	.697	.668	.689	.663	.685	.658	.616	.701
		.8	.534	.570	.521	.565	.517	.559	.440	.614

Anmerkung: Vorpalte 1 gibt die Verteilung an, Vorpalte 2 den Stichprobenumfang N , Vorpalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-25. $|\overline{CP} - .95|$

			Perc ($d_s^{S_1}$)		BCA ($d_s^{S_1}$)		Perc-t ($d_s^{S_1}$)	
			1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	.036	.041	.013	.014	.015	.009
		.4	.033	.038	.011	.015	.009	.009
		.8	.038	.045	.017	.017	.017	.010
	60	.0	.012	.013	.004	.005	.005	.004
		.4	.011	.016	.005	.007	.004	.007
		.8	.013	.016	.007	.006	.006	.004
Gleich	20	.0	.026	.022	.007	.007	.003	.014
		.4	.022	.018	.008	.006	.006	.016
		.8	.018	.017	.012	.010	.014	.020
	60	.0	.010	.007	.004	.003	.001	.006
		.4	.008	.005	.003	.001	.001	.006
		.8	.004	.004	.004	.004	.003	.007
g-h-1	20	.0	.096	.124	.052	.058	.068	.077
		.4	.116	.132	.066	.065	.080	.084
		.8	.167	.161	.098	.086	.106	.095
	60	.0	.061	.078	.039	.044	.060	.068
		.4	.076	.088	.053	.051	.075	.078
		.8	.114	.101	.068	.059	.092	.086
g-h-2	20	.0	.065	.058	.022	.027	.031	.035
		.4	.085	.069	.038	.035	.056	.044
		.8	.133	.102	.075	.060	.090	.057
	60	.0	.035	.028	.015	.018	.026	.027
		.4	.048	.045	.025	.026	.038	.037
		.8	.065	.060	.039	.039	.051	.049
g-h-3	20	.0	.114	.128	.057	.059	.068	.067
		.4	.145	.147	.078	.074	.091	.079
		.8	.217	.178	.128	.097	.139	.095
	60	.0	.088	.092	.050	.052	.072	.075
		.4	.108	.114	.066	.065	.091	.091
		.8	.164	.136	.102	.080	.135	.107
g-h-4	20	.0	.028	.017	.003	.008	.010	.019
		.4	.031	.018	.007	.007	.008	.018
		.8	.047	.019	.025	.013	.030	.030
	60	.0	.009	.006	.001	.003	.003	.010
		.4	.007	.006	.001	.002	.004	.008
		.8	.011	.004	.005	.003	.011	.012
Hyper	20	.0	.029	.016	.008	.010	.010	.018
		.4	.028	.016	.008	.009	.005	.016
		.8	.033	.014	.017	.010	.012	.018
	60	.0	.008	.005	.003	.003	.003	.011
		.4	.008	.005	.002	.002	.001	.012
		.8	.006	.008	.002	.003	.004	.009

Anmerkung: Vorpalte 1 gibt die Verteilung an, Vorpalte 2 den Stichprobenumfang N , Vorpalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-26. Mittlere Intervalllänge

			Perc ($d_s^{S_1}$)		BCA ($d_s^{S_1}$)		Perc-t ($d_s^{S_1}$)	
			1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	1.532	1.531	1.475	1.469	1.409	1.395
		.4	1.269	1.370	1.207	1.305	1.159	1.241
		.8	.908	1.190	.831	1.121	.806	1.068
	60	.0	.820	.817	.812	.808	.801	.797
		.4	.672	.732	.662	.722	.655	.712
		.8	.475	.637	.463	.626	.459	.618
Gleich	20	.0	1.444	1.479	1.406	1.429	1.357	1.363
		.4	1.152	1.308	1.115	1.257	1.084	1.202
		.8	.775	1.115	.729	1.062	.714	1.018
	60	.0	.773	.778	.769	.772	.762	.762
		.4	.616	.688	.611	.682	.607	.673
		.8	.402	.585	.396	.578	.393	.571
g-h-1	20	.0	1.865	1.765	1.753	1.648	1.600	1.513
		.4	1.592	1.591	1.467	1.473	1.336	1.351
		.8	1.235	1.425	1.096	1.307	.985	1.189
	60	.0	1.013	.970	.988	.950	.946	.914
		.4	.867	.889	.841	.871	.802	.835
		.8	.691	.806	.674	.789	.627	.752
g-h-2	20	.0	1.537	1.301	1.518	1.247	1.316	1.155
		.4	1.355	1.151	1.296	1.098	1.133	1.023
		.8	1.165	1.013	1.046	.960	.901	.885
	60	.0	.776	.660	.772	.651	.732	.636
		.4	.688	.590	.676	.584	.644	.569
		.8	.608	.526	.588	.522	.553	.504
g-h-3	20	.0	1.831	1.636	1.755	1.545	1.574	1.418
		.4	1.570	1.471	1.460	1.373	1.308	1.260
		.8	1.280	1.324	1.134	1.223	.995	1.110
	60	.0	.971	.881	.959	.872	.901	.829
		.4	.851	.804	.836	.797	.782	.756
		.8	.722	.739	.709	.731	.645	.687
g-h-4	20	.0	1.354	1.233	1.326	1.181	1.226	1.133
		.4	1.149	1.079	1.094	1.024	1.029	.992
		.8	.923	.908	.823	.851	.776	.830
	60	.0	.688	.628	.687	.623	.673	.617
		.4	.578	.546	.572	.539	.564	.536
		.8	.443	.455	.428	.446	.425	.446
Hyper	20	.0	1.474	1.444	1.443	1.404	1.370	1.336
		.4	1.230	1.276	1.188	1.235	1.146	1.179
		.8	.943	1.090	.884	1.048	.868	1.002
	60	.0	.795	.772	.790	.767	.778	.756
		.4	.656	.681	.650	.675	.644	.666
		.8	.506	.580	.496	.574	.494	.567

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-27. $|\overline{CP} - .95|$

			Perc (\hat{d}^{S_1})		BCA (\hat{d}^{S_1})		Perc-t (\hat{d}^{S_1})		NC ($d_s^{S_1}$)	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	.031	.034	.015	.015	.018	.013	.005	.002
		.4	.027	.031	.011	.015	.014	.013	.003	.003
		.8	.027	.034	.023	.019	.029	.016	.003	.001
	60	.0	.011	.011	.005	.006	.005	.005	.003	.002
		.4	.010	.014	.006	.008	.006	.009	.002	.002
		.8	.009	.013	.007	.007	.009	.006	.004	.001
Gleich	20	.0	.023	.020	.010	.008	.006	.015	.006	.011
		.4	.020	.018	.013	.008	.011	.018	.011	.010
		.8	.015	.015	.021	.014	.035	.029	.021	.016
	60	.0	.009	.007	.004	.003	.002	.007	.008	.009
		.4	.007	.006	.004	.002	.003	.007	.012	.013
		.8	.004	.004	.005	.002	.009	.009	.025	.016
g-h-1	20	.0	.081	.108	.046	.051	.067	.074	.050	.054
		.4	.100	.115	.057	.057	.077	.082	.065	.064
		.8	.142	.139	.083	.074	.101	.092	.115	.080
	60	.0	.055	.071	.036	.040	.059	.066	.079	.083
		.4	.069	.079	.049	.048	.072	.077	.105	.097
		.8	.102	.093	.061	.054	.088	.083	.164	.116
g-h-2	20	.0	.051	.041	.021	.026	.034	.043	.016	.033
		.4	.068	.050	.035	.032	.060	.051	.019	.032
		.8	.111	.075	.068	.051	.096	.069	.066	.030
	60	.0	.030	.021	.015	.017	.027	.029	.015	.031
		.4	.042	.036	.024	.024	.039	.039	.025	.030
		.8	.057	.049	.038	.035	.052	.051	.077	.029
g-h-3	20	.0	.097	.110	.050	.052	.066	.065	.039	.032
		.4	.123	.124	.068	.063	.084	.075	.061	.042
		.8	.187	.151	.111	.084	.131	.091	.139	.061
	60	.0	.079	.084	.046	.048	.069	.072	.074	.060
		.4	.099	.104	.061	.060	.088	.087	.112	.080
		.8	.152	.125	.094	.073	.129	.103	.212	.106
g-h-4	20	.0	.020	.011	.006	.011	.010	.023	.027	.033
		.4	.022	.013	.012	.010	.012	.032	.022	.032
		.8	.035	.012	.031	.018	.046	.047	.011	.035
	60	.0	.008	.005	.002	.003	.003	.012	.027	.032
		.4	.005	.007	.002	.004	.004	.013	.026	.034
		.8	.009	.003	.008	.007	.016	.019	.018	.036
Hyper	20	.0	.027	.016	.011	.009	.011	.018	.009	.012
		.4	.024	.017	.010	.010	.009	.019	.003	.012
		.8	.027	.014	.023	.013	.019	.027	.018	.015
	60	.0	.007	.004	.004	.001	.004	.009	.006	.013
		.4	.007	.006	.004	.003	.003	.013	.004	.014
		.8	.006	.007	.004	.004	.005	.013	.011	.018

Anmerkung: Vospalte 1 gibt die Verteilung an, Vospalte 2 den Stichprobenumfang N , Vospalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-28. Mittlere Intervalllänge

			Perc (\hat{d}^{S_1})		BCA (\hat{d}^{S_1})		Perc-t (\hat{d}^{S_1})		NC ($d_s^{S_1}$)	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	1.471	1.469	1.416	1.411	1.353	1.339	1.418	1.411
		.4	1.218	1.315	1.158	1.253	1.113	1.192	1.173	1.263
		.8	.872	1.142	.798	1.076	.773	1.026	.832	1.099
	60	.0	.810	.807	.802	.798	.791	.787	.802	.800
		.4	.664	.722	.654	.713	.647	.703	.658	.717
		.8	.469	.628	.457	.618	.453	.610	.465	.623
Gleich	20	.0	1.386	1.420	1.350	1.372	1.303	1.309	1.399	1.400
		.4	1.106	1.256	1.071	1.207	1.041	1.153	1.149	1.253
		.8	.744	1.070	.700	1.019	.685	.977	.815	1.090
	60	.0	.763	.768	.759	.763	.752	.753	.798	.798
		.4	.608	.680	.603	.673	.599	.665	.654	.715
		.8	.397	.578	.391	.570	.388	.564	.462	.621
g-h-1	20	.0	1.790	1.695	1.683	1.582	1.536	1.453	1.523	1.475
		.4	1.529	1.527	1.408	1.414	1.282	1.297	1.256	1.307
		.8	1.185	1.368	1.052	1.255	.946	1.141	.900	1.142
	60	.0	1.000	.958	.975	.938	.934	.902	.833	.817
		.4	.856	.877	.830	.860	.792	.824	.680	.728
		.8	.682	.796	.665	.779	.619	.743	.483	.633
g-h-2	20	.0	1.475	1.249	1.457	1.198	1.264	1.109	1.483	1.428
		.4	1.301	1.105	1.244	1.054	1.088	.982	1.229	1.273
		.8	1.118	.972	1.004	.922	.865	.850	.883	1.111
	60	.0	.766	.652	.762	.643	.722	.628	.817	.804
		.4	.680	.582	.667	.577	.636	.562	.669	.719
		.8	.600	.519	.580	.516	.546	.498	.475	.625
g-h-3	20	.0	1.758	1.570	1.685	1.483	1.511	1.361	1.563	1.482
		.4	1.508	1.412	1.401	1.318	1.255	1.209	1.274	1.310
		.8	1.229	1.271	1.089	1.174	.956	1.066	.919	1.144
	60	.0	.959	.870	.947	.861	.889	.819	.848	.822
		.4	.841	.794	.825	.787	.772	.746	.691	.730
		.8	.712	.730	.700	.722	.636	.679	.491	.635
g-h-4	20	.0	1.300	1.184	1.273	1.134	1.177	1.088	1.420	1.399
		.4	1.103	1.036	1.051	.983	.987	.953	1.167	1.252
		.8	.886	.872	.790	.817	.745	.797	.833	1.088
	60	.0	.679	.620	.678	.615	.664	.609	.801	.797
		.4	.570	.539	.565	.532	.557	.530	.656	.714
		.8	.438	.449	.423	.440	.420	.440	.464	.620
Hyper	20	.0	1.415	1.386	1.385	1.348	1.316	1.282	1.409	1.396
		.4	1.181	1.225	1.140	1.186	1.101	1.132	1.161	1.250
		.8	.906	1.047	.849	1.006	.833	.962	.827	1.087
	60	.0	.785	.762	.780	.757	.768	.746	.800	.797
		.4	.648	.672	.641	.667	.636	.658	.656	.714
		.8	.499	.573	.489	.567	.487	.560	.466	.620

Anmerkung: Vorpalte 1 gibt die Verteilung an, Vorpalte 2 den Stichprobenumfang N , Vorpalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-29. $|\overline{CP} - .95|$

			Perc ($d_s^{S_2}$)		BCA ($d_s^{S_2}$)		Perc-t ($d_s^{S_2}$)	
			1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	.037	.027	.014	.008	.016	.011
		.4	.032	.028	.012	.009	.011	.011
		.8	.041	.035	.020	.014	.017	.013
	60	.0	.014	.011	.006	.004	.008	.004
		.4	.011	.012	.004	.006	.005	.006
		.8	.014	.012	.008	.004	.007	.005
Gleich	20	.0	.023	.021	.007	.003	.002	.008
		.4	.022	.017	.008	.004	.004	.008
		.8	.018	.017	.013	.007	.014	.009
	60	.0	.009	.005	.002	.002	.002	.003
		.4	.006	.009	.002	.003	.001	.002
		.8	.006	.005	.005	.003	.006	.004
g-h-1	20	.0	.099	.081	.054	.050	.072	.065
		.4	.119	.094	.064	.059	.082	.073
		.8	.169	.130	.096	.079	.109	.088
	60	.0	.064	.056	.043	.044	.063	.057
		.4	.080	.066	.053	.048	.076	.063
		.8	.118	.086	.071	.054	.098	.075
g-h-2	20	.0	.084	.062	.053	.041	.079	.053
		.4	.098	.072	.063	.049	.087	.056
		.8	.132	.086	.084	.054	.098	.056
	60	.0	.039	.030	.024	.023	.040	.027
		.4	.046	.041	.029	.031	.045	.032
		.8	.065	.050	.041	.032	.053	.034
g-h-3	20	.0	.131	.100	.084	.067	.106	.086
		.4	.161	.119	.098	.078	.120	.094
		.8	.221	.153	.135	.095	.150	.099
	60	.0	.094	.080	.063	.064	.097	.081
		.4	.120	.094	.080	.073	.114	.091
		.8	.168	.122	.109	.081	.145	.101
g-h-4	20	.0	.039	.030	.016	.009	.022	.009
		.4	.041	.031	.023	.012	.030	.010
		.8	.047	.024	.032	.012	.038	.017
	60	.0	.013	.011	.002	.003	.005	.003
		.4	.009	.011	.002	.004	.005	.004
		.8	.014	.005	.004	.003	.010	.008
Hyper	20	.0	.046	.025	.025	.009	.032	.024
		.4	.048	.024	.029	.009	.032	.020
		.8	.058	.025	.036	.014	.035	.018
	60	.0	.014	.010	.006	.003	.008	.004
		.4	.014	.008	.006	.004	.008	.005
		.8	.015	.010	.008	.006	.009	.009

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-30. Mittlere Intervalllänge

			Perc ($d_s^{S_2}$)		BCA ($d_s^{S_2}$)		Perc-t ($d_s^{S_2}$)	
			1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	1.536	1.550	1.480	1.498	1.414	1.439
		.4	1.267	1.390	1.204	1.335	1.158	1.286
		.8	.907	1.202	.830	1.137	.805	1.093
	60	.0	.821	.825	.812	.817	.801	.808
		.4	.674	.739	.664	.730	.657	.723
		.8	.475	.639	.463	.628	.460	.622
Gleich	20	.0	1.445	1.420	1.407	1.392	1.359	1.364
		.4	1.155	1.260	1.118	1.231	1.087	1.210
		.8	.775	1.095	.729	1.055	.714	1.026
	60	.0	.774	.771	.769	.768	.762	.765
		.4	.617	.680	.612	.676	.608	.674
		.8	.402	.581	.396	.576	.394	.572
g-h-1	20	.0	1.875	1.978	1.763	1.887	1.611	1.733
		.4	1.595	1.780	1.469	1.668	1.342	1.542
		.8	1.235	1.515	1.098	1.385	.984	1.278
	60	.0	1.010	1.049	.984	1.025	.941	.982
		.4	.870	.956	.845	.929	.804	.894
		.8	.692	.837	.674	.814	.628	.782
g-h-2	20	.0	2.349	2.154	2.075	1.988	1.873	1.906
		.4	1.929	1.754	1.693	1.625	1.531	1.595
		.8	1.408	1.271	1.214	1.166	1.070	1.136
	60	.0	1.186	1.091	1.122	1.050	1.084	1.039
		.4	.996	.908	.943	.877	.911	.874
		.8	.738	.668	.697	.648	.667	.641
g-h-3	20	.0	2.136	2.131	1.966	2.005	1.804	1.914
		.4	1.782	1.856	1.613	1.729	1.475	1.649
		.8	1.357	1.471	1.196	1.340	1.057	1.257
	60	.0	1.177	1.158	1.130	1.125	1.078	1.101
		.4	1.006	1.015	.966	.982	.915	.960
		.8	.778	.822	.754	.799	.694	.769
g-h-4	20	.0	1.986	1.799	1.796	1.690	1.669	1.633
		.4	1.612	1.465	1.447	1.386	1.354	1.362
		.8	1.127	1.078	.982	1.011	.920	1.000
	60	.0	.971	.896	.947	.884	.930	.878
		.4	.785	.742	.764	.733	.752	.731
		.8	.538	.543	.517	.535	.511	.535
Hyper	20	.0	2.057	1.589	1.912	1.533	1.700	1.473
		.4	1.747	1.395	1.596	1.346	1.424	1.297
		.8	1.328	1.193	1.164	1.142	1.034	1.088
	60	.0	1.055	.868	1.015	.856	.988	.847
		.4	.881	.758	.842	.747	.822	.740
		.8	.665	.644	.623	.633	.610	.624

Anmerkung: Vorpalte 1 gibt die Verteilung an, Vorpalte 2 den Stichprobenumfang N , Vorpalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-31. $|\overline{CP} - .95|$

			Perc (\hat{d}^{S_2})		BCA (\hat{d}^{S_2})		Perc-t (\hat{d}^{S_2})		NC ($d_s^{S_2}$)	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	.032	.023	.015	.011	.019	.015	.003	.004
		.4	.025	.022	.013	.012	.015	.015	.003	.004
		.8	.029	.028	.023	.016	.028	.018	.003	.001
	60	.0	.013	.010	.006	.005	.008	.006	.004	.002
		.4	.008	.011	.005	.006	.006	.008	.003	.003
		.8	.010	.010	.011	.006	.010	.007	.003	.002
Gleich	20	.0	.022	.019	.010	.002	.005	.006	.006	.007
		.4	.018	.016	.012	.005	.011	.005	.010	.007
		.8	.016	.015	.023	.011	.034	.013	.021	.012
	60	.0	.008	.004	.003	.003	.002	.003	.010	.010
		.4	.006	.009	.003	.005	.002	.004	.014	.009
		.8	.006	.004	.008	.004	.012	.005	.024	.016
g-h-1	20	.0	.086	.069	.049	.047	.070	.065	.052	.051
		.4	.100	.081	.055	.053	.079	.071	.067	.061
		.8	.143	.112	.083	.069	.106	.087	.115	.079
	60	.0	.058	.051	.041	.042	.062	.055	.079	.080
		.4	.073	.059	.050	.044	.072	.062	.107	.096
		.8	.106	.078	.065	.050	.094	.072	.169	.116
g-h-2	20	.0	.075	.056	.051	.041	.083	.055	.105	.079
		.4	.086	.063	.059	.046	.090	.057	.110	.068
		.8	.114	.069	.076	.048	.100	.058	.132	.025
	60	.0	.036	.028	.023	.023	.042	.027	.114	.083
		.4	.042	.037	.028	.030	.046	.032	.125	.071
		.8	.057	.042	.040	.030	.053	.035	.152	.030
g-h-3	20	.0	.117	.088	.075	.061	.103	.083	.104	.078
		.4	.141	.102	.087	.071	.113	.090	.121	.083
		.8	.193	.132	.117	.082	.140	.094	.174	.076
	60	.0	.089	.073	.060	.061	.095	.078	.154	.129
		.4	.111	.087	.075	.068	.110	.089	.188	.134
		.8	.155	.112	.101	.074	.140	.097	.255	.128
g-h-4	20	.0	.035	.029	.017	.010	.028	.012	.043	.033
		.4	.037	.028	.025	.014	.038	.013	.041	.014
		.8	.041	.020	.035	.016	.051	.024	.028	.018
	60	.0	.012	.012	.003	.003	.007	.003	.038	.024
		.4	.008	.010	.002	.004	.008	.004	.031	.004
		.8	.011	.003	.006	.005	.015	.012	.021	.026
Hyper	20	.0	.042	.021	.025	.009	.037	.025	.059	.015
		.4	.043	.020	.029	.010	.037	.022	.061	.009
		.8	.050	.019	.037	.014	.043	.023	.069	.004
	60	.0	.013	.008	.006	.004	.011	.004	.054	.014
		.4	.013	.007	.007	.005	.010	.005	.057	.007
		.8	.014	.008	.010	.007	.012	.011	.063	.005

Anmerkung: Vospalte 1 gibt die Verteilung an, Vospalte 2 den Stichprobenumfang N , Vospalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-32. Mittlere Intervalllänge

			Perc (\hat{d}^{S_2})		BCA (\hat{d}^{S_2})		Perc-t (\hat{d}^{S_2})		NC ($d_s^{S_2}$)	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	1.474	1.488	1.421	1.438	1.358	1.381	1.423	1.433
		.4	1.216	1.335	1.156	1.282	1.111	1.235	1.171	1.284
		.8	.871	1.154	.797	1.091	.773	1.049	.833	1.111
	60	.0	.810	.815	.802	.807	.791	.798	.801	.806
		.4	.665	.730	.656	.721	.649	.714	.659	.722
		.8	.469	.631	.457	.620	.454	.614	.465	.625
Gleich	20	.0	1.387	1.363	1.351	1.336	1.304	1.309	1.401	1.408
		.4	1.109	1.209	1.073	1.182	1.043	1.162	1.149	1.263
		.8	.744	1.051	.700	1.012	.685	.985	.815	1.096
	60	.0	.764	.761	.759	.758	.752	.755	.798	.802
		.4	.609	.671	.605	.667	.600	.666	.655	.717
		.8	.397	.574	.391	.568	.389	.565	.462	.622
g-h-1	20	.0	1.800	1.899	1.692	1.811	1.547	1.664	1.533	1.570
		.4	1.531	1.709	1.410	1.601	1.289	1.480	1.256	1.392
		.8	1.185	1.454	1.054	1.329	.944	1.227	.899	1.184
	60	.0	.997	1.036	.972	1.012	.929	.969	.831	.846
		.4	.859	.944	.834	.917	.794	.882	.682	.754
		.8	.683	.827	.665	.804	.620	.772	.483	.647
g-h-2	20	.0	2.255	2.067	1.992	1.909	1.798	1.830	1.520	1.533
		.4	1.852	1.684	1.625	1.560	1.470	1.531	1.250	1.359
		.8	1.351	1.220	1.165	1.120	1.027	1.091	.891	1.150
	60	.0	1.171	1.077	1.108	1.037	1.071	1.026	.820	.827
		.4	.983	.897	.931	.866	.899	.863	.674	.736
		.8	.728	.660	.688	.639	.658	.633	.477	.634
g-h-3	20	.0	2.051	2.046	1.887	1.924	1.732	1.837	1.585	1.611
		.4	1.711	1.782	1.549	1.660	1.416	1.583	1.290	1.425
		.8	1.303	1.412	1.149	1.287	1.014	1.207	.924	1.194
	60	.0	1.162	1.143	1.116	1.110	1.064	1.087	.849	.867
		.4	.993	1.002	.954	.970	.904	.948	.693	.766
		.8	.768	.811	.744	.789	.685	.759	.492	.654
g-h-4	20	.0	1.906	1.727	1.724	1.622	1.602	1.568	1.434	1.437
		.4	1.547	1.406	1.389	1.330	1.300	1.308	1.177	1.285
		.8	1.082	1.035	.943	.971	.883	.960	.838	1.104
	60	.0	.958	.884	.935	.873	.918	.867	.804	.804
		.4	.775	.732	.754	.724	.742	.722	.658	.719
		.8	.531	.536	.510	.528	.504	.528	.465	.623
Hyper	20	.0	1.975	1.526	1.836	1.472	1.632	1.414	1.445	1.444
		.4	1.677	1.340	1.532	1.292	1.367	1.245	1.196	1.301
		.8	1.275	1.145	1.117	1.096	.993	1.045	.856	1.129
	60	.0	1.041	.857	1.002	.845	.975	.836	.807	.806
		.4	.870	.749	.831	.738	.812	.730	.662	.723
		.8	.656	.636	.616	.625	.603	.616	.471	.628

Anmerkung: Vorpalte 1 gibt die Verteilung an, Vorpalte 2 den Stichprobenumfang N , Vorpalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-33. $|\overline{CP} - .95|$

		Perc ($d_{D,S,R}$)		BCA ($d_{D,S,R}$)		Perc-t ($d_{D,S,R}$)		Perc ($\hat{d}_{D,R}$)		BCA ($\hat{d}_{D,R}$)		Perc-t ($\hat{d}_{D,R}$)		NC ($d_{D,S,R}$)		
		1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	
Normal	20	.0	.007	.005	.011	.007	.016	.011	.009	.007	.009	.005	.019	.016	.004	.003
		.4	.007	.008	.010	.007	.015	.016	.011	.009	.010	.006	.023	.023	.011	.007
		.8	.014	.010	.024	.007	.029	.021	.015	.014	.026	.005	.044	.032	.008	.013
	60	.0	.004	.002	.005	.003	.008	.011	.002	.002	.006	.003	.010	.013	.001	.005
		.4	.005	.006	.004	.004	.008	.011	.007	.007	.003	.003	.009	.014	.006	.008
		.8	.009	.007	.012	.002	.015	.017	.009	.008	.013	.002	.018	.021	.005	.014
Gleich	20	.0	.017	.011	.010	.006	.005	.007	.011	.008	.008	.005	.007	.010	.007	.007
		.4	.017	.006	.015	.005	.009	.010	.007	.006	.014	.005	.016	.017	.012	.006
		.8	.027	.005	.035	.008	.029	.022	.009	.008	.032	.009	.041	.032	.011	.005
	60	.0	.006	.002	.003	.004	.004	.013	.004	.001	.003	.004	.006	.015	.005	.002
		.4	.005	.005	.006	.001	.005	.010	.003	.005	.007	.002	.006	.013	.007	.005
		.8	.008	.004	.013	.003	.013	.016	.003	.006	.014	.002	.015	.020	.006	.003
g-h-1	20	.0	.014	.010	.008	.008	.024	.025	.013	.012	.007	.005	.027	.031	.005	.015
		.4	.019	.015	.009	.006	.025	.035	.018	.018	.010	.004	.033	.042	.006	.024
		.8	.029	.019	.025	.008	.038	.042	.021	.022	.033	.007	.053	.052	.017	.043
	60	.0	.006	.007	.005	.004	.014	.017	.007	.008	.004	.003	.015	.019	.007	.019
		.4	.012	.011	.003	.002	.012	.019	.010	.012	.004	.002	.013	.022	.003	.025
		.8	.021	.015	.006	.003	.014	.026	.016	.016	.008	.004	.017	.031	.008	.045
g-h-2	20	.0	.010	.014	.010	.006	.028	.017	.011	.018	.010	.003	.034	.025	.009	.027
		.4	.010	.013	.007	.009	.038	.030	.013	.019	.009	.007	.046	.040	.020	.026
		.8	.014	.014	.021	.019	.062	.040	.018	.022	.028	.016	.076	.057	.047	.027
	60	.0	.004	.012	.006	.003	.021	.011	.006	.012	.005	.003	.021	.014	.015	.029
		.4	.008	.015	.002	.004	.020	.018	.008	.017	.002	.004	.022	.022	.021	.030
		.8	.012	.015	.006	.006	.032	.030	.011	.018	.007	.005	.033	.037	.048	.032
g-h-3	20	.0	.013	.015	.006	.006	.023	.019	.013	.016	.005	.006	.027	.025	.007	.006
		.4	.022	.017	.008	.008	.026	.029	.019	.019	.012	.007	.033	.037	.009	.011
		.8	.030	.021	.026	.011	.043	.037	.022	.024	.035	.011	.059	.049	.033	.026
	60	.0	.008	.010	.005	.003	.013	.012	.009	.010	.003	.003	.014	.014	.008	.010
		.4	.015	.015	.004	.004	.010	.015	.014	.015	.004	.004	.010	.017	.008	.014
		.8	.022	.018	.007	.004	.019	.022	.016	.019	.009	.004	.020	.025	.018	.032
g-h-4	20	.0	.010	.006	.011	.006	.019	.017	.005	.012	.008	.004	.023	.019	.006	.029
		.4	.007	.009	.013	.008	.031	.015	.004	.016	.011	.007	.037	.024	.012	.031
		.8	.004	.011	.027	.015	.052	.028	.009	.020	.029	.012	.067	.041	.028	.030
	60	.0	.002	.008	.002	.003	.012	.011	.005	.010	.002	.002	.012	.012	.005	.030
		.4	.002	.009	.004	.004	.017	.013	.003	.012	.003	.003	.019	.016	.013	.034
		.8	.006	.013	.011	.005	.029	.018	.007	.017	.012	.004	.033	.021	.032	.034
Hyper	20	.0	.011	.008	.021	.009	.033	.028	.011	.009	.025	.011	.042	.034	.030	.004
		.4	.010	.011	.027	.013	.034	.032	.008	.010	.033	.017	.047	.042	.021	.005
		.8	.020	.012	.043	.020	.072	.048	.016	.012	.056	.028	.093	.066	.049	.009
	60	.0	.006	.003	.019	.009	.024	.041	.004	.004	.021	.012	.028	.047	.037	.008
		.4	.009	.005	.022	.017	.031	.049	.006	.006	.026	.020	.037	.056	.028	.012
		.8	.017	.006	.045	.027	.071	.059	.012	.010	.054	.031	.081	.067	.060	.022

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-34. Mittlere Intervalllänge

			Perc ($d_{D,s,R}$)		BCA ($d_{D,s,R}$)		Perc-t ($d_{D,s,R}$)		Perc ($\hat{d}_{D,R}$)		BCA ($\hat{d}_{D,R}$)		Perc-t ($\hat{d}_{D,R}$)		NC ($d_{D,s,R}$)	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	1.886	2.067	1.867	2.042	1.673	1.774	1.754	1.923	1.737	1.900	1.556	1.650	1.623	1.634
		.4	1.604	1.960	1.594	1.935	1.428	1.665	1.492	1.823	1.483	1.800	1.328	1.549	1.383	1.500
		.8	1.249	1.918	1.243	1.901	1.108	1.580	1.162	1.784	1.157	1.769	1.030	1.470	1.042	1.338
	60	.0	.940	.979	.936	.976	.910	.942	.919	.958	.916	.955	.891	.921	.899	.901
		.4	.796	.918	.793	.914	.772	.880	.779	.898	.775	.894	.756	.861	.767	.825
		.8	.605	.863	.603	.859	.589	.819	.592	.845	.590	.841	.577	.802	.577	.731
Gleich	20	.0	2.261	2.489	2.191	2.411	1.954	2.065	2.103	2.316	2.038	2.243	1.818	1.921	1.941	1.949
		.4	1.851	2.313	1.775	2.229	1.599	1.886	1.722	2.151	1.651	2.074	1.487	1.754	1.620	1.766
		.8	1.391	2.168	1.304	2.100	1.163	1.699	1.294	2.017	1.213	1.954	1.082	1.580	1.173	1.545
	60	.0	1.110	1.141	1.099	1.131	1.069	1.093	1.086	1.117	1.076	1.107	1.046	1.069	1.077	1.077
		.4	.913	1.046	.902	1.035	.880	.997	.894	1.024	.882	1.012	.861	.975	.898	.974
		.8	.665	.936	.651	.925	.636	.883	.651	.916	.637	.905	.623	.864	.646	.849
g-h-1	20	.0	1.519	1.693	1.549	1.716	1.414	1.521	1.413	1.575	1.441	1.596	1.315	1.415	1.250	1.267
		.4	1.332	1.650	1.374	1.671	1.271	1.479	1.239	1.535	1.278	1.554	1.183	1.376	1.077	1.171
		.8	1.122	1.731	1.167	1.756	1.132	1.524	1.044	1.610	1.085	1.634	1.053	1.418	.852	1.079
	60	.0	.753	.803	.757	.806	.739	.781	.737	.786	.740	.789	.723	.764	.692	.693
		.4	.650	.781	.656	.783	.643	.758	.636	.764	.642	.766	.629	.742	.597	.640
		.8	.539	.786	.547	.788	.548	.757	.527	.769	.535	.771	.536	.740	.477	.583
g-h-2	20	.0	1.786	1.554	1.763	1.550	1.593	1.337	1.661	1.446	1.640	1.442	1.482	1.243	1.428	1.351
		.4	1.603	1.484	1.570	1.463	1.439	1.251	1.491	1.380	1.461	1.361	1.339	1.163	1.218	1.233
		.8	1.424	1.505	1.393	1.472	1.306	1.214	1.325	1.400	1.296	1.369	1.215	1.129	.955	1.109
	60	.0	.883	.723	.880	.721	.858	.698	.864	.708	.861	.705	.839	.682	.788	.748
		.4	.785	.676	.780	.671	.764	.648	.768	.662	.763	.657	.748	.634	.672	.678
		.8	.686	.662	.683	.654	.679	.626	.671	.647	.668	.640	.664	.613	.523	.602
g-h-3	20	.0	1.490	1.514	1.520	1.546	1.400	1.361	1.386	1.409	1.414	1.438	1.303	1.266	1.205	1.180
		.4	1.307	1.478	1.352	1.509	1.271	1.326	1.216	1.375	1.258	1.404	1.182	1.234	1.036	1.091
		.8	1.123	1.558	1.163	1.585	1.165	1.367	1.044	1.449	1.082	1.474	1.083	1.272	.827	1.005
	60	.0	.735	.712	.739	.716	.723	.693	.719	.697	.723	.700	.707	.678	.665	.648
		.4	.640	.691	.645	.693	.636	.670	.626	.676	.631	.678	.622	.655	.574	.596
		.8	.544	.701	.551	.703	.561	.674	.533	.686	.539	.688	.549	.659	.463	.543
g-h-4	20	.0	2.105	1.869	2.048	1.829	1.825	1.584	1.958	1.739	1.905	1.702	1.698	1.474	1.737	1.666
		.4	1.850	1.746	1.774	1.688	1.583	1.444	1.721	1.625	1.651	1.570	1.473	1.343	1.462	1.506
		.8	1.564	1.714	1.484	1.642	1.294	1.331	1.455	1.595	1.381	1.527	1.204	1.238	1.092	1.323
	60	.0	1.036	.857	1.028	.852	.999	.827	1.014	.839	1.006	.833	.978	.809	.961	.925
		.4	.898	.777	.887	.769	.863	.746	.878	.761	.868	.752	.845	.730	.806	.833
		.8	.726	.717	.714	.707	.695	.678	.711	.702	.699	.692	.680	.664	.596	.726
Hyper	20	.0	2.197	3.266	2.184	3.135	1.991	2.412	2.044	3.038	2.031	2.916	1.852	2.244	1.721	2.161
		.4	1.863	3.113	1.859	2.970	1.702	2.249	1.733	2.896	1.729	2.763	1.584	2.093	1.484	1.998
		.8	1.682	3.164	1.635	2.970	1.536	2.029	1.564	2.953	1.520	2.775	1.432	1.890	1.171	1.774
	60	.0	1.140	1.269	1.125	1.258	1.108	1.209	1.115	1.241	1.101	1.231	1.084	1.183	.958	1.195
		.4	.944	1.173	.930	1.162	.923	1.113	.924	1.148	.910	1.137	.903	1.089	.824	1.100
		.8	.854	1.068	.847	1.056	.827	1.005	.836	1.045	.828	1.033	.809	.984	.640	.974

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-35. $|\overline{CP} - .95|$

			Perc ($d'_{s,R}$)		BCA ($d'_{s,R}$)		Perc-t ₁ ($d'_{s,R}$)		Perc-t ₂ ($d'_{s,R}$)	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	.007	.006	.017	.011	.022	.019	.016	.007
		.4	.010	.007	.011	.010	.019	.023	.014	.010
		.8	.017	.008	.009	.003	.020	.017	.016	.009
	60	.0	.006	.002	.009	.004	.010	.011	.007	.010
		.4	.007	.005	.005	.004	.008	.011	.005	.010
		.8	.009	.003	.006	.003	.010	.013	.008	.012
Gleich	20	.0	.011	.009	.018	.009	.010	.005	.006	.012
		.4	.005	.006	.019	.007	.011	.007	.005	.013
		.8	.010	.005	.020	.005	.021	.013	.016	.016
	60	.0	.004	.002	.007	.003	.002	.012	.005	.012
		.4	.002	.004	.007	.002	.003	.010	.004	.011
		.8	.006	.002	.009	.002	.008	.012	.006	.012
g-h-1	20	.0	.014	.012	.016	.012	.028	.034	.026	.022
		.4	.017	.015	.010	.010	.025	.038	.024	.026
		.8	.024	.016	.009	.005	.024	.028	.019	.018
	60	.0	.010	.008	.007	.005	.017	.020	.015	.014
		.4	.011	.010	.005	.003	.012	.020	.011	.015
		.8	.014	.010	.003	.002	.014	.019	.012	.017
g-h-2	20	.0	.011	.015	.015	.010	.032	.021	.029	.019
		.4	.009	.013	.015	.011	.041	.025	.039	.022
		.8	.007	.011	.017	.010	.051	.032	.046	.027
	60	.0	.007	.011	.007	.004	.023	.012	.021	.011
		.4	.004	.014	.007	.006	.026	.016	.025	.015
		.8	.006	.012	.005	.003	.029	.021	.027	.019
g-h-3	20	.0	.015	.015	.014	.008	.029	.026	.026	.016
		.4	.020	.017	.008	.009	.027	.029	.026	.017
		.8	.023	.019	.007	.008	.030	.025	.025	.018
	60	.0	.010	.011	.006	.003	.017	.014	.015	.012
		.4	.014	.013	.003	.004	.014	.014	.012	.011
		.8	.015	.012	.004	.005	.017	.016	.014	.013
g-h-4	20	.0	.006	.009	.018	.009	.025	.016	.021	.018
		.4	.004	.010	.016	.011	.033	.013	.031	.014
		.8	.005	.007	.014	.010	.043	.021	.040	.019
	60	.0	.003	.010	.005	.005	.015	.012	.013	.011
		.4	.004	.008	.005	.007	.016	.011	.015	.010
		.8	.005	.008	.005	.005	.027	.014	.026	.013
Hyper	20	.0	.010	.008	.033	.010	.036	.027	.031	.031
		.4	.010	.011	.036	.015	.036	.027	.033	.031
		.8	.015	.012	.037	.018	.044	.034	.039	.034
	60	.0	.006	.003	.025	.009	.025	.041	.029	.041
		.4	.007	.004	.024	.015	.026	.046	.030	.047
		.8	.012	.005	.029	.016	.040	.050	.038	.050

Anmerkung: Vorsepalte 1 gibt die Verteilung an, Vorsepalte 2 den Stichprobenumfang N , Vorsepalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-36. Mittlere Intervalllänge

			Perc ($d'_{s,R}$)		BCA ($d'_{s,R}$)		Perc-t ₁ ($d'_{s,R}$)		Perc-t ₂ ($d'_{s,R}$)	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	1.723	1.951	1.749	1.954	1.610	1.733	1.603	1.775
		.4	1.495	1.812	1.533	1.818	1.410	1.608	1.389	1.644
		.8	1.262	1.703	1.293	1.701	1.186	1.483	1.168	1.515
	60	.0	.890	.957	.892	.957	.875	.927	.877	.935
		.4	.762	.882	.765	.882	.750	.854	.748	.861
		.8	.607	.804	.610	.802	.596	.775	.594	.780
Gleich	20	.0	2.088	2.367	2.076	2.328	1.899	2.029	1.918	2.101
		.4	1.729	2.148	1.721	2.108	1.581	1.838	1.580	1.900
		.8	1.348	1.954	1.318	1.904	1.196	1.632	1.191	1.675
	60	.0	1.062	1.122	1.059	1.116	1.038	1.081	1.046	1.090
		.4	.874	1.015	.871	1.008	.854	.977	.857	.985
		.8	.633	.893	.627	.884	.612	.856	.613	.860
g-h-1	20	.0	1.365	1.575	1.419	1.614	1.337	1.469	1.289	1.480
		.4	1.242	1.498	1.310	1.539	1.240	1.398	1.184	1.408
		.8	1.182	1.494	1.237	1.528	1.207	1.377	1.169	1.394
	60	.0	.702	.779	.709	.784	.699	.764	.693	.769
		.4	.624	.735	.633	.740	.624	.721	.616	.725
		.8	.564	.705	.572	.709	.565	.689	.562	.692
g-h-2	20	.0	1.636	1.467	1.651	1.489	1.533	1.301	1.495	1.334
		.4	1.512	1.372	1.514	1.384	1.418	1.210	1.369	1.242
		.8	1.445	1.318	1.405	1.299	1.369	1.130	1.332	1.156
	60	.0	.836	.704	.837	.704	.824	.683	.818	.689
		.4	.756	.647	.755	.645	.741	.627	.734	.633
		.8	.683	.600	.678	.596	.667	.579	.663	.582
g-h-3	20	.0	1.338	1.413	1.391	1.461	1.322	1.316	1.265	1.328
		.4	1.225	1.341	1.293	1.395	1.236	1.254	1.173	1.266
		.8	1.196	1.339	1.243	1.372	1.242	1.231	1.202	1.247
	60	.0	.683	.690	.691	.695	.682	.676	.674	.680
		.4	.614	.649	.623	.654	.617	.636	.607	.640
		.8	.570	.624	.577	.628	.573	.609	.570	.612
g-h-4	20	.0	1.955	1.777	1.945	1.770	1.777	1.550	1.762	1.598
		.4	1.782	1.636	1.752	1.613	1.589	1.412	1.564	1.459
		.8	1.651	1.525	1.563	1.470	1.413	1.262	1.391	1.299
	60	.0	.991	.839	.989	.836	.969	.814	.969	.820
		.4	.872	.754	.866	.750	.845	.732	.842	.739
		.8	.729	.666	.717	.658	.695	.642	.693	.647
Hyper	20	.0	2.042	3.152	2.037	3.064	1.941	2.386	1.930	2.517
		.4	1.785	2.899	1.792	2.821	1.743	2.216	1.722	2.330
		.8	1.511	2.713	1.504	2.646	1.545	1.995	1.535	2.073
	60	.0	1.098	1.255	1.083	1.248	1.081	1.200	1.081	1.208
		.4	.945	1.153	.932	1.146	.937	1.103	.936	1.110
		.8	.804	1.037	.796	1.029	.807	.985	.807	.989

Anmerkung: Vorpalte 1 gibt die Verteilung an, Vorpalte 2 den Stichprobenumfang N , Vorpalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-37. $|\overline{CP} - .95|$

			Perc ($\hat{d}'_{cl,R}$)		BCA ($\hat{d}'_{cl,R}$)		Perc-t ($\hat{d}'_{cl,R}$)		NC ₁ ($d'_{s,R}$)	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	.007	.009	.015	.008	.022	.023	.009	.003
		.4	.011	.008	.011	.007	.022	.027	.015	.002
		.8	.019	.011	.011	.004	.029	.026	.011	.003
	60	.0	.005	.003	.008	.003	.011	.013	.007	.003
		.4	.008	.005	.004	.004	.008	.012	.011	.003
		.8	.009	.005	.005	.002	.012	.015	.008	.004
Gleich	20	.0	.007	.008	.018	.008	.011	.009	.009	.009
		.4	.007	.006	.017	.005	.015	.011	.017	.007
		.8	.019	.007	.017	.004	.032	.022	.019	.007
	60	.0	.003	.001	.006	.004	.002	.013	.009	.005
		.4	.002	.004	.007	.003	.004	.011	.013	.007
		.8	.008	.005	.008	.002	.011	.014	.018	.008
g-h-1	20	.0	.010	.012	.015	.009	.028	.038	.005	.009
		.4	.013	.016	.009	.008	.026	.041	.007	.015
		.8	.022	.017	.009	.006	.028	.038	.011	.018
	60	.0	.008	.008	.007	.005	.017	.021	.005	.013
		.4	.010	.009	.005	.003	.013	.021	.005	.015
		.8	.012	.010	.003	.002	.015	.020	.017	.023
g-h-2	20	.0	.012	.020	.015	.009	.035	.028	.003	.029
		.4	.012	.018	.012	.006	.044	.034	.012	.028
		.8	.014	.020	.014	.007	.055	.040	.032	.028
	60	.0	.007	.012	.006	.003	.023	.015	.006	.032
		.4	.005	.015	.006	.006	.025	.019	.015	.031
		.8	.007	.014	.004	.003	.029	.026	.040	.033
g-h-3	20	.0	.010	.014	.013	.008	.030	.030	.004	.006
		.4	.016	.018	.007	.009	.028	.035	.004	.007
		.8	.022	.020	.009	.007	.033	.034	.021	.009
	60	.0	.008	.011	.006	.004	.016	.017	.004	.009
		.4	.012	.013	.002	.004	.012	.016	.004	.009
		.8	.013	.013	.005	.004	.016	.018	.023	.014
g-h-4	20	.0	.004	.013	.016	.005	.026	.018	.005	.029
		.4	.007	.017	.016	.009	.036	.020	.009	.031
		.8	.011	.017	.013	.007	.053	.031	.021	.032
	60	.0	.004	.011	.004	.005	.015	.013	.004	.032
		.4	.006	.011	.005	.004	.017	.014	.004	.034
		.8	.007	.012	.005	.003	.030	.018	.022	.036
Hyper	20	.0	.011	.008	.037	.013	.045	.032	.026	.005
		.4	.010	.009	.039	.016	.043	.035	.025	.005
		.8	.018	.011	.043	.022	.057	.049	.042	.005
	60	.0	.003	.005	.027	.014	.029	.046	.032	.007
		.4	.002	.006	.028	.018	.032	.054	.030	.010
		.8	.007	.008	.032	.020	.044	.057	.044	.009

Anmerkung: Vorpalte 1 gibt die Verteilung an, Vorpalte 2 den Stichprobenumfang N , Vorpalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-38. Mittlere Intervalllänge

			Perc ($\hat{d}'_{c1,R}$)		BCA ($\hat{d}'_{c1,R}$)		Perc-t ($\hat{d}'_{c1,R}$)		NC ₁ ($d'_{s,R}$)	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	1.603	1.814	1.627	1.818	1.498	1.612	1.604	1.628
		.4	1.391	1.685	1.426	1.691	1.312	1.495	1.393	1.494
		.8	1.174	1.584	1.202	1.582	1.103	1.380	1.091	1.331
	60	.0	.871	.937	.873	.936	.856	.907	.896	.900
		.4	.745	.863	.749	.863	.734	.836	.769	.824
		.8	.594	.787	.597	.785	.583	.758	.586	.730
Gleich	20	.0	1.942	2.202	1.931	2.165	1.767	1.887	1.920	1.942
		.4	1.609	1.998	1.601	1.961	1.470	1.710	1.625	1.760
		.8	1.254	1.818	1.226	1.771	1.112	1.518	1.210	1.544
	60	.0	1.040	1.098	1.036	1.092	1.016	1.057	1.073	1.076
		.4	.855	.993	.852	.986	.836	.956	.900	.974
		.8	.619	.874	.613	.865	.599	.837	.651	.849
g-h-1	20	.0	1.270	1.465	1.320	1.501	1.244	1.366	1.233	1.260
		.4	1.155	1.393	1.218	1.432	1.154	1.300	1.091	1.165
		.8	1.099	1.390	1.150	1.422	1.123	1.281	.911	1.069
	60	.0	.687	.762	.694	.767	.684	.747	.689	.692
		.4	.610	.719	.620	.724	.610	.705	.600	.639
		.8	.552	.690	.559	.693	.553	.674	.488	.582
g-h-2	20	.0	1.522	1.365	1.536	1.385	1.426	1.210	1.412	1.346
		.4	1.407	1.277	1.408	1.287	1.319	1.126	1.236	1.228
		.8	1.345	1.226	1.307	1.208	1.274	1.051	1.013	1.099
	60	.0	.818	.689	.819	.689	.806	.669	.786	.747
		.4	.739	.633	.739	.631	.725	.614	.675	.677
		.8	.669	.587	.663	.583	.653	.566	.535	.601
g-h-3	20	.0	1.244	1.315	1.294	1.359	1.229	1.224	1.190	1.175
		.4	1.139	1.248	1.203	1.298	1.150	1.166	1.052	1.086
		.8	1.113	1.246	1.156	1.276	1.155	1.145	.891	.995
	60	.0	.668	.675	.676	.680	.668	.661	.663	.647
		.4	.601	.635	.609	.640	.603	.622	.577	.595
		.8	.558	.611	.564	.614	.561	.596	.476	.543
g-h-4	20	.0	1.818	1.653	1.809	1.647	1.653	1.442	1.719	1.657
		.4	1.658	1.522	1.629	1.501	1.478	1.314	1.477	1.502
		.8	1.536	1.419	1.454	1.368	1.314	1.174	1.151	1.315
	60	.0	.970	.820	.967	.818	.948	.796	.958	.924
		.4	.853	.738	.848	.734	.826	.717	.808	.833
		.8	.713	.652	.702	.644	.680	.629	.606	.725
Hyper	20	.0	1.900	2.932	1.894	2.850	1.806	2.219	1.702	2.156
		.4	1.661	2.697	1.667	2.625	1.621	2.061	1.492	2.001
		.8	1.406	2.523	1.399	2.461	1.438	1.857	1.185	1.801
	60	.0	1.074	1.228	1.060	1.221	1.058	1.174	.955	1.194
		.4	.925	1.128	.912	1.121	.917	1.079	.825	1.105
		.8	.786	1.015	.779	1.006	.789	.964	.644	.989

Anmerkung: Vorpalte 1 gibt die Verteilung an, Vorpalte 2 den Stichprobenumfang N , Vorpalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-39. $|\overline{CP} - .95|$

			Perc ($\hat{d}'_{c2,R}$)		BCA ($\hat{d}'_{c2,R}$)		Perc-t ($\hat{d}'_{c2,R}$)		NC ₂ ($d'_{s,R}$)	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	.008	.008	.016	.008	.015	.012	.009	.013
		.4	.012	.008	.011	.007	.017	.016	.003	.013
		.8	.020	.011	.012	.004	.024	.019	.006	.009
	60	.0	.005	.003	.009	.003	.008	.012	.009	.008
		.4	.008	.006	.004	.004	.006	.012	.005	.011
		.8	.010	.004	.005	.002	.010	.014	.008	.009
Gleich	20	.0	.009	.009	.018	.008	.006	.015	.009	.011
		.4	.007	.006	.017	.005	.009	.017	.004	.009
		.8	.018	.007	.017	.004	.026	.026	.009	.007
	60	.0	.004	.001	.006	.004	.005	.014	.004	.002
		.4	.002	.004	.007	.003	.005	.012	.003	.004
		.8	.008	.005	.008	.002	.009	.014	.009	.005
g-h-1	20	.0	.013	.013	.016	.010	.026	.028	.018	.025
		.4	.016	.016	.009	.008	.026	.032	.016	.029
		.8	.023	.017	.009	.006	.025	.028	.029	.027
	60	.0	.009	.008	.007	.004	.015	.016	.020	.024
		.4	.011	.010	.005	.004	.012	.016	.024	.026
		.8	.013	.010	.003	.002	.014	.019	.038	.030
g-h-2	20	.0	.012	.019	.015	.009	.031	.024	.023	.026
		.4	.012	.018	.013	.007	.042	.032	.034	.023
		.8	.013	.020	.015	.006	.051	.035	.057	.024
	60	.0	.007	.012	.007	.004	.021	.013	.027	.029
		.4	.005	.015	.007	.006	.024	.018	.043	.028
		.8	.008	.014	.004	.004	.028	.024	.065	.031
g-h-3	20	.0	.013	.015	.014	.008	.028	.022	.020	.011
		.4	.018	.018	.008	.009	.028	.026	.025	.012
		.8	.023	.020	.010	.007	.030	.027	.043	.014
	60	.0	.009	.010	.006	.003	.015	.014	.022	.011
		.4	.013	.014	.002	.004	.012	.013	.030	.013
		.8	.014	.013	.005	.004	.014	.015	.046	.018
g-h-4	20	.0	.005	.013	.018	.005	.023	.018	.013	.028
		.4	.006	.016	.015	.009	.034	.019	.022	.031
		.8	.011	.017	.013	.007	.049	.029	.041	.031
	60	.0	.004	.011	.005	.004	.013	.013	.012	.031
		.4	.005	.010	.005	.005	.016	.013	.020	.032
		.8	.007	.012	.005	.003	.029	.017	.040	.036
Hyper	20	.0	.009	.008	.035	.013	.037	.037	.048	.004
		.4	.009	.010	.038	.016	.039	.038	.042	.006
		.8	.019	.011	.042	.022	.051	.050	.052	.004
	60	.0	.004	.004	.026	.014	.031	.046	.052	.007
		.4	.004	.005	.027	.017	.033	.054	.052	.011
		.8	.009	.008	.031	.019	.043	.057	.058	.009

Anmerkung: Vorpalte 1 gibt die Verteilung an, Vorpalte 2 den Stichprobenumfang N , Vorpalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-40. Mittlere Intervalllänge

			Perc ($\hat{d}'_{c2,R}$)		BCA ($\hat{d}'_{c2,R}$)		Perc-t ($\hat{d}'_{c2,R}$)		NC ₂ ($d'_{s,R}$)	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	1.650	1.856	1.681	1.860	1.544	1.686	1.502	1.577
		.4	1.431	1.722	1.474	1.729	1.338	1.560	1.292	1.446
		.8	1.204	1.610	1.239	1.609	1.119	1.428	1.017	1.304
	60	.0	.880	.942	.882	.942	.867	.920	.840	.878
		.4	.753	.868	.757	.868	.740	.847	.711	.803
		.8	.599	.790	.602	.788	.586	.765	.546	.718
Gleich	20	.0	1.998	2.250	1.993	2.214	1.846	1.990	1.824	1.896
		.4	1.654	2.039	1.654	2.004	1.520	1.798	1.530	1.718
		.8	1.284	1.845	1.263	1.798	1.139	1.576	1.142	1.520
	60	.0	1.050	1.104	1.047	1.097	1.034	1.072	1.024	1.057
		.4	.863	.998	.860	.991	.847	.969	.848	.956
		.8	.624	.877	.618	.868	.604	.844	.615	.839
g-h-1	20	.0	1.308	1.501	1.365	1.539	1.245	1.410	1.119	1.201
		.4	1.189	1.426	1.261	1.467	1.144	1.340	.979	1.112
		.8	1.127	1.413	1.184	1.447	1.120	1.315	.838	1.041
	60	.0	.694	.768	.701	.772	.685	.757	.621	.666
		.4	.616	.724	.626	.729	.609	.714	.533	.615
		.8	.556	.692	.564	.696	.554	.679	.448	.570
g-h-2	20	.0	1.563	1.392	1.585	1.417	1.442	1.263	1.307	1.306
		.4	1.445	1.303	1.455	1.315	1.320	1.176	1.130	1.190
		.8	1.376	1.246	1.343	1.228	1.273	1.089	.939	1.077
	60	.0	.826	.692	.827	.693	.809	.677	.724	.729
		.4	.746	.637	.746	.635	.725	.622	.613	.660
		.8	.673	.589	.668	.585	.654	.572	.495	.592
g-h-3	20	.0	1.282	1.346	1.337	1.393	1.223	1.264	1.073	1.122
		.4	1.173	1.277	1.244	1.329	1.134	1.203	.939	1.038
		.8	1.140	1.266	1.189	1.298	1.150	1.176	.820	.971
	60	.0	.675	.680	.683	.684	.667	.670	.592	.623
		.4	.607	.639	.616	.644	.601	.630	.509	.573
		.8	.562	.613	.569	.617	.562	.601	.437	.532
g-h-4	20	.0	1.864	1.683	1.863	1.679	1.694	1.510	1.624	1.623
		.4	1.700	1.551	1.680	1.531	1.504	1.378	1.380	1.469
		.8	1.570	1.440	1.494	1.389	1.328	1.222	1.080	1.295
	60	.0	.979	.824	.977	.822	.958	.806	.905	.909
		.4	.861	.741	.856	.737	.832	.726	.753	.818
		.8	.719	.654	.707	.646	.683	.634	.568	.717
Hyper	20	.0	1.952	2.989	1.953	2.905	1.852	2.369	1.597	2.123
		.4	1.706	2.747	1.720	2.675	1.652	2.192	1.390	1.970
		.8	1.434	2.555	1.432	2.492	1.458	1.941	1.125	1.786
	60	.0	1.083	1.232	1.069	1.225	1.067	1.185	.896	1.182
		.4	.932	1.132	.919	1.124	.924	1.089	.767	1.093
		.8	.791	1.016	.784	1.008	.794	.969	.611	.984

Anmerkung: Vorpalte 1 gibt die Verteilung an, Vorpalte 2 den Stichprobenumfang N , Vorpalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-41. $|\overline{CP} - .95|$

			Perc ($d_{s,R}^{S_1}$)		BCA ($d_{s,R}^{S_1}$)		Perc-t ($d_{s,R}^{S_1}$)	
			1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	.009	.006	.009	.007	.017	.008
		.4	.011	.008	.007	.006	.016	.011
		.8	.016	.008	.009	.003	.020	.011
	60	.0	.005	.002	.006	.002	.009	.011
		.4	.008	.006	.006	.003	.012	.010
		.8	.011	.004	.003	.003	.012	.013
Gleich	20	.0	.009	.008	.013	.007	.006	.009
		.4	.005	.006	.016	.005	.008	.011
		.8	.011	.005	.018	.005	.018	.017
	60	.0	.003	.002	.005	.005	.004	.013
		.4	.002	.005	.006	.002	.005	.011
		.8	.008	.003	.008	.002	.010	.013
g-h-1	20	.0	.013	.011	.007	.007	.024	.021
		.4	.017	.014	.006	.008	.026	.026
		.8	.023	.016	.010	.006	.024	.023
	60	.0	.009	.008	.004	.003	.016	.016
		.4	.012	.012	.003	.002	.016	.015
		.8	.016	.012	.004	.003	.019	.019
g-h-2	20	.0	.017	.014	.010	.007	.017	.022
		.4	.014	.012	.011	.009	.027	.026
		.8	.008	.011	.017	.008	.043	.028
	60	.0	.010	.011	.006	.004	.017	.014
		.4	.008	.014	.006	.006	.023	.017
		.8	.007	.012	.005	.004	.038	.022
g-h-3	20	.0	.018	.014	.009	.006	.021	.016
		.4	.023	.015	.007	.007	.023	.019
		.8	.023	.019	.010	.008	.029	.023
	60	.0	.012	.012	.004	.001	.014	.013
		.4	.016	.013	.004	.005	.016	.014
		.8	.016	.013	.004	.004	.022	.015
g-h-4	20	.0	.010	.009	.006	.006	.011	.016
		.4	.007	.009	.010	.008	.016	.015
		.8	.005	.008	.013	.010	.035	.022
	60	.0	.010	.010	.002	.004	.009	.013
		.4	.007	.009	.003	.005	.013	.013
		.8	.006	.009	.006	.004	.034	.014
Hyper	20	.0	.017	.009	.023	.010	.028	.032
		.4	.016	.011	.030	.013	.031	.031
		.8	.020	.012	.034	.017	.041	.037
	60	.0	.010	.003	.032	.011	.042	.045
		.4	.012	.004	.031	.016	.044	.051
		.8	.018	.006	.035	.018	.058	.053

Anmerkung: Vospalte 1 gibt die Verteilung an, Vospalte 2 den Stichprobenumfang N , Vospalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-42. Mittlere Intervalllänge

			Perc ($d_{s,R}^{S_1}$)		BCA ($d_{s,R}^{S_1}$)		Perc-t ($d_{s,R}^{S_1}$)	
			1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	2.290	2.270	2.327	2.270	1.901	1.860
		.4	2.011	2.103	2.059	2.105	1.690	1.724
		.8	1.645	1.897	1.705	1.903	1.392	1.559
	60	.0	1.015	1.009	1.013	1.006	.969	.963
		.4	.882	.931	.882	.928	.843	.888
		.8	.697	.837	.697	.834	.666	.798
Gleich	20	.0	2.635	2.681	2.595	2.617	2.136	2.129
		.4	2.214	2.431	2.182	2.371	1.799	1.928
		.8	1.699	2.136	1.673	2.076	1.352	1.690
	60	.0	1.157	1.161	1.148	1.152	1.103	1.105
		.4	.965	1.053	.956	1.043	.919	1.001
		.8	.700	.918	.690	.907	.662	.871
g-h-1	20	.0	1.956	1.920	2.041	1.964	1.695	1.628
		.4	1.777	1.803	1.863	1.845	1.577	1.537
		.8	1.560	1.689	1.644	1.739	1.448	1.462
	60	.0	.855	.844	.861	.847	.825	.812
		.4	.770	.795	.776	.799	.747	.767
		.8	.660	.742	.665	.744	.645	.717
g-h-2	20	.0	2.037	1.730	2.070	1.738	1.637	1.377
		.4	1.894	1.611	1.888	1.604	1.569	1.285
		.8	1.750	1.477	1.688	1.454	1.515	1.185
	60	.0	.854	.742	.850	.738	.806	.704
		.4	.797	.685	.789	.679	.756	.651
		.8	.730	.627	.720	.620	.698	.598
g-h-3	20	.0	1.846	1.704	1.946	1.765	1.595	1.444
		.4	1.705	1.608	1.805	1.676	1.517	1.372
		.8	1.544	1.509	1.625	1.560	1.449	1.305
	60	.0	.798	.745	.804	.749	.768	.717
		.4	.733	.702	.739	.705	.711	.676
		.8	.651	.657	.656	.660	.639	.634
g-h-4	20	.0	2.401	2.112	2.394	2.080	1.851	1.620
		.4	2.189	1.918	2.123	1.866	1.703	1.479
		.8	1.976	1.701	1.839	1.629	1.536	1.310
	60	.0	.964	.868	.957	.862	.905	.823
		.4	.878	.784	.866	.776	.825	.747
		.8	.758	.689	.741	.679	.707	.657
Hyper	20	.0	2.449	3.962	2.482	3.787	2.024	2.521
		.4	2.197	3.615	2.220	3.462	1.850	2.342
		.8	1.776	3.203	1.806	3.099	1.563	2.030
	60	.0	1.129	1.277	1.113	1.269	1.066	1.209
		.4	.994	1.175	.986	1.167	.941	1.114
		.8	.832	1.051	.823	1.041	.795	.992

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-43. $|\overline{CP} - .95|$

			Perc ($\hat{d}_R^{S_1}$)		BCA ($\hat{d}_R^{S_1}$)		Perc-t ($\hat{d}_R^{S_1}$)		NC ($d_{s,R}^{S_1}$)	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	.012	.009	.008	.004	.024	.014	.008	.007
		.4	.014	.009	.007	.005	.023	.019	.006	.010
		.8	.021	.013	.010	.003	.032	.021	.008	.008
	60	.0	.005	.003	.005	.002	.012	.013	.008	.007
		.4	.009	.006	.004	.003	.014	.014	.007	.010
		.8	.013	.006	.004	.002	.016	.016	.009	.009
Gleich	20	.0	.009	.009	.012	.005	.011	.012	.002	.007
		.4	.010	.007	.014	.004	.016	.017	.006	.005
		.8	.021	.008	.015	.004	.033	.028	.011	.006
	60	.0	.004	.001	.004	.005	.005	.015	.002	.002
		.4	.003	.005	.005	.003	.007	.013	.004	.005
		.8	.010	.005	.008	.002	.014	.017	.009	.005
g-h-1	20	.0	.015	.013	.007	.007	.033	.031	.025	.022
		.4	.019	.017	.006	.007	.033	.035	.024	.027
		.8	.027	.018	.011	.007	.034	.035	.035	.027
	60	.0	.009	.008	.003	.003	.017	.018	.024	.026
		.4	.012	.011	.002	.002	.019	.019	.029	.027
		.8	.016	.011	.003	.002	.022	.020	.042	.031
g-h-2	20	.0	.020	.021	.010	.004	.028	.032	.016	.029
		.4	.020	.018	.008	.005	.036	.037	.017	.026
		.8	.016	.020	.014	.007	.054	.043	.037	.026
	60	.0	.013	.013	.006	.004	.021	.018	.017	.031
		.4	.011	.016	.004	.005	.025	.021	.022	.030
		.8	.010	.015	.005	.002	.040	.029	.046	.031
g-h-3	20	.0	.019	.016	.008	.006	.030	.025	.015	.010
		.4	.024	.019	.008	.007	.033	.030	.020	.012
		.8	.026	.021	.010	.006	.037	.031	.040	.015
	60	.0	.012	.012	.003	.002	.016	.016	.015	.013
		.4	.015	.014	.003	.004	.018	.015	.025	.014
		.8	.017	.014	.005	.004	.023	.018	.044	.019
g-h-4	20	.0	.017	.015	.006	.002	.014	.020	.023	.032
		.4	.016	.017	.007	.006	.023	.025	.015	.031
		.8	.014	.018	.010	.006	.049	.033	.024	.031
	60	.0	.013	.013	.002	.003	.010	.014	.024	.034
		.4	.011	.012	.003	.004	.016	.016	.012	.034
		.8	.007	.013	.004	.003	.037	.019	.026	.035
Hyper	20	.0	.016	.009	.024	.012	.042	.038	.011	.004
		.4	.017	.011	.030	.016	.044	.042	.016	.005
		.8	.018	.011	.038	.021	.058	.054	.035	.004
	60	.0	.008	.005	.035	.015	.049	.051	.026	.009
		.4	.009	.005	.034	.019	.050	.058	.029	.013
		.8	.013	.009	.038	.021	.065	.062	.047	.013

Anmerkung: Vospalte 1 gibt die Verteilung an, Vospalte 2 den Stichprobenumfang N , Vospalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-44. Mittlere Intervalllänge

			Perc ($\hat{d}_R^{S_1}$)		BCA ($\hat{d}_R^{S_1}$)		Perc-t ($\hat{d}_R^{S_1}$)		NC ($d_{s,R}^{S_1}$)	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	2.131	2.111	2.165	2.111	1.768	1.730	1.681	1.653
		.4	1.871	1.956	1.915	1.958	1.572	1.604	1.457	1.514
		.8	1.530	1.765	1.586	1.771	1.295	1.450	1.124	1.342
	60	.0	.993	.987	.992	.985	.948	.942	.908	.903
		.4	.863	.911	.863	.908	.825	.869	.778	.827
		.8	.682	.819	.682	.816	.651	.781	.591	.732
Gleich	20	.0	2.451	2.494	2.414	2.434	1.987	1.981	1.991	1.964
		.4	2.059	2.261	2.030	2.206	1.674	1.794	1.680	1.777
		.8	1.580	1.987	1.556	1.932	1.258	1.572	1.238	1.553
	60	.0	1.132	1.136	1.123	1.128	1.079	1.081	1.083	1.079
		.4	.944	1.030	.936	1.021	.900	.979	.908	.976
		.8	.685	.898	.675	.888	.647	.853	.655	.850
g-h-1	20	.0	1.819	1.786	1.899	1.827	1.577	1.514	1.307	1.286
		.4	1.653	1.677	1.733	1.716	1.467	1.430	1.150	1.185
		.8	1.451	1.571	1.529	1.617	1.347	1.360	.941	1.080
	60	.0	.837	.825	.842	.829	.808	.794	.700	.696
		.4	.754	.778	.759	.782	.731	.750	.609	.642
		.8	.646	.726	.651	.728	.631	.701	.492	.584
g-h-2	20	.0	1.894	1.609	1.925	1.617	1.523	1.281	1.506	1.378
		.4	1.761	1.498	1.756	1.492	1.460	1.195	1.311	1.250
		.8	1.628	1.374	1.571	1.353	1.410	1.103	1.048	1.109
	60	.0	.835	.726	.832	.723	.788	.689	.801	.752
		.4	.780	.670	.772	.664	.739	.637	.686	.680
		.8	.714	.614	.705	.607	.683	.585	.540	.603
g-h-3	20	.0	1.717	1.585	1.810	1.641	1.484	1.343	1.258	1.199
		.4	1.586	1.496	1.679	1.559	1.411	1.276	1.108	1.104
		.8	1.436	1.404	1.511	1.451	1.348	1.214	.920	1.005
	60	.0	.780	.729	.787	.733	.752	.701	.674	.651
		.4	.717	.687	.723	.690	.696	.661	.586	.598
		.8	.637	.643	.642	.645	.625	.620	.480	.544
g-h-4	20	.0	2.234	1.964	2.227	1.935	1.722	1.507	1.836	1.694
		.4	2.036	1.785	1.975	1.736	1.584	1.376	1.558	1.525
		.8	1.838	1.582	1.710	1.515	1.429	1.218	1.187	1.325
	60	.0	.943	.849	.936	.844	.885	.805	.975	.930
		.4	.859	.767	.848	.759	.807	.730	.820	.836
		.8	.742	.674	.725	.664	.692	.643	.610	.726
Hyper	20	.0	2.278	3.686	2.309	3.523	1.883	2.345	1.794	2.169
		.4	2.044	3.363	2.065	3.221	1.722	2.178	1.571	2.011
		.8	1.652	2.980	1.680	2.883	1.455	1.891	1.222	1.803
	60	.0	1.104	1.250	1.089	1.242	1.043	1.183	.971	1.192
		.4	.972	1.150	.965	1.141	.920	1.090	.839	1.102
		.8	.814	1.028	.805	1.019	.778	.971	.651	.986

Anmerkung: Vorpalte 1 gibt die Verteilung an, Vorpalte 2 den Stichprobenumfang N , Vorpalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-45. $|\overline{CP} - .95|$

			Perc ($d_{s,R}^{S_2}$)		BCA ($d_{s,R}^{S_2}$)		Perc-t ($d_{s,R}^{S_2}$)	
			1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	.009	.012	.009	.006	.016	.016
		.4	.012	.010	.008	.008	.015	.016
		.8	.018	.013	.010	.005	.022	.014
	60	.0	.005	.005	.007	.004	.011	.006
		.4	.009	.006	.002	.006	.007	.009
		.8	.012	.006	.004	.004	.013	.009
Gleich	20	.0	.007	.008	.013	.010	.005	.012
		.4	.006	.005	.013	.009	.008	.012
		.8	.012	.005	.020	.005	.021	.007
	60	.0	.004	.001	.007	.003	.005	.002
		.4	.003	.003	.006	.004	.003	.003
		.8	.006	.004	.007	.001	.008	.004
g-h-1	20	.0	.014	.016	.007	.010	.026	.023
		.4	.019	.021	.007	.011	.025	.021
		.8	.024	.020	.009	.009	.026	.023
	60	.0	.010	.011	.005	.003	.018	.011
		.4	.010	.010	.004	.005	.015	.015
		.8	.016	.013	.004	.003	.018	.016
g-h-2	20	.0	.007	.002	.010	.007	.055	.036
		.4	.003	.006	.008	.005	.056	.026
		.8	.004	.010	.020	.011	.060	.026
	60	.0	.002	.003	.003	.002	.034	.016
		.4	.002	.003	.003	.005	.036	.018
		.8	.003	.009	.004	.004	.040	.019
g-h-3	20	.0	.009	.013	.007	.008	.035	.026
		.4	.016	.019	.006	.010	.032	.024
		.8	.023	.021	.007	.010	.032	.024
	60	.0	.006	.008	.005	.005	.022	.017
		.4	.012	.010	.002	.004	.020	.017
		.8	.015	.014	.003	.004	.022	.015
g-h-4	20	.0	.008	.005	.008	.006	.041	.026
		.4	.007	.007	.010	.010	.046	.024
		.8	.003	.007	.014	.012	.056	.018
	60	.0	.003	.003	.003	.004	.021	.011
		.4	.003	.004	.003	.004	.023	.009
		.8	.004	.008	.005	.004	.034	.011
Hyper	20	.0	.008	.013	.040	.054	.117	.064
		.4	.009	.013	.045	.063	.121	.065
		.8	.019	.016	.052	.088	.125	.071
	60	.0	.007	.012	.020	.033	.085	.034
		.4	.006	.013	.022	.037	.087	.034
		.8	.012	.016	.030	.044	.105	.055

Anmerkung: Vorpalte 1 gibt die Verteilung an, Vorpalte 2 den Stichprobenumfang N , Vorpalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-46. Mittlere Intervalllänge

			Perc ($d_{s,R}^{S_2}$)		BCA ($d_{s,R}^{S_2}$)		Perc-t ($d_{s,R}^{S_2}$)	
			1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	2.272	2.312	2.304	2.362	1.897	2.013
		.4	2.014	2.144	2.061	2.202	1.681	1.894
		.8	1.648	1.937	1.708	1.994	1.386	1.701
	60	.0	1.016	1.022	1.015	1.024	.971	.990
		.4	.885	.943	.885	.943	.845	.916
		.8	.696	.841	.696	.841	.665	.814
Gleich	20	.0	2.621	2.585	2.578	2.541	2.132	2.232
		.4	2.225	2.355	2.192	2.318	1.804	2.056
		.8	1.710	2.113	1.684	2.075	1.355	1.801
	60	.0	1.158	1.153	1.150	1.145	1.106	1.122
		.4	.969	1.038	.960	1.029	.923	1.011
		.8	.701	.910	.692	.901	.663	.879
g-h-1	20	.0	1.946	2.013	2.028	2.114	1.692	1.844
		.4	1.792	1.892	1.886	2.012	1.584	1.785
		.8	1.567	1.748	1.655	1.847	1.441	1.643
	60	.0	.859	.867	.863	.875	.828	.847
		.4	.771	.809	.777	.816	.747	.794
		.8	.661	.749	.666	.755	.646	.737
g-h-2	20	.0	2.830	2.443	2.667	2.343	2.300	2.193
		.4	2.440	2.085	2.305	2.031	2.018	1.928
		.8	1.997	1.703	1.904	1.696	1.743	1.547
	60	.0	1.196	1.018	1.174	1.004	1.117	.988
		.4	1.034	.877	1.015	.866	.968	.861
		.8	.833	.716	.818	.708	.792	.703
g-h-3	20	.0	2.032	1.951	2.069	2.024	1.792	1.831
		.4	1.841	1.803	1.905	1.901	1.654	1.740
		.8	1.599	1.600	1.667	1.689	1.515	1.512
	60	.0	.892	.843	.892	.847	.864	.830
		.4	.795	.769	.797	.774	.774	.760
		.8	.678	.686	.680	.691	.666	.677
g-h-4	20	.0	3.602	3.051	3.349	2.881	2.624	2.561
		.4	3.004	2.522	2.804	2.417	2.231	2.172
		.8	2.340	1.971	2.188	1.910	1.784	1.687
	60	.0	1.403	1.200	1.374	1.179	1.279	1.147
		.4	1.177	1.014	1.150	.998	1.075	.986
		.8	.893	.794	.871	.781	.820	.775
Hyper	20	.0	2.778	2.697	2.632	2.593	3.047	2.888
		.4	2.439	2.480	2.322	2.357	2.619	2.712
		.8	1.947	2.252	1.827	2.078	2.245	2.428
	60	.0	1.587	1.571	1.559	1.505	1.677	1.482
		.4	1.397	1.445	1.374	1.367	1.516	1.368
		.8	1.139	1.293	1.107	1.194	1.305	1.214

Anmerkung: Vorderspalte 1 gibt die Verteilung an, Vorderspalte 2 den Stichprobenumfang N , Vorderspalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-47. $|\overline{CP} - .95|$

			Perc ($\hat{d}_R^{S_2}$)		BCA ($\hat{d}_R^{S_2}$)		Perc-t ($\hat{d}_R^{S_2}$)		NC ($d_{s,R}^{S_2}$)	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	.011	.013	.009	.006	.023	.023	.006	.009
		.4	.016	.014	.008	.008	.023	.022	.005	.010
		.8	.023	.016	.009	.005	.035	.021	.008	.009
	60	.0	.006	.006	.007	.003	.013	.006	.009	.005
		.4	.010	.007	.002	.006	.010	.011	.006	.010
		.8	.012	.007	.004	.003	.016	.011	.011	.009
Gleich	20	.0	.008	.010	.011	.007	.010	.014	.003	.007
		.4	.010	.011	.011	.006	.015	.013	.007	.005
		.8	.021	.011	.017	.003	.036	.012	.011	.006
	60	.0	.004	.003	.005	.002	.006	.002	.001	.004
		.4	.004	.005	.005	.005	.006	.003	.003	.002
		.8	.009	.006	.007	.001	.013	.006	.008	.006
g-h-1	20	.0	.014	.018	.007	.010	.032	.030	.024	.026
		.4	.020	.021	.009	.010	.033	.026	.025	.026
		.8	.026	.022	.009	.010	.037	.032	.034	.030
	60	.0	.011	.012	.004	.003	.020	.013	.027	.024
		.4	.011	.009	.004	.005	.017	.017	.032	.030
		.8	.017	.013	.004	.004	.021	.018	.040	.032
g-h-2	20	.0	.006	.005	.009	.006	.063	.038	.077	.049
		.4	.007	.011	.006	.004	.062	.030	.077	.034
		.8	.011	.017	.016	.009	.067	.033	.080	.020
	60	.0	.002	.004	.004	.002	.035	.018	.086	.050
		.4	.003	.006	.003	.005	.037	.018	.087	.038
		.8	.006	.011	.004	.003	.042	.020	.092	.022
g-h-3	20	.0	.012	.015	.008	.008	.041	.032	.043	.036
		.4	.019	.020	.007	.010	.039	.028	.046	.029
		.8	.026	.023	.007	.009	.041	.032	.052	.023
	60	.0	.008	.008	.005	.004	.024	.017	.049	.039
		.4	.013	.011	.002	.004	.021	.016	.051	.035
		.8	.016	.015	.003	.004	.023	.016	.058	.024
g-h-4	20	.0	.005	.004	.007	.005	.052	.030	.065	.040
		.4	.005	.009	.010	.008	.054	.027	.060	.027
		.8	.010	.015	.013	.011	.068	.025	.062	.013
	60	.0	.001	.003	.003	.003	.023	.013	.066	.042
		.4	.004	.004	.003	.004	.026	.011	.062	.020
		.8	.007	.010	.003	.004	.039	.012	.067	.014
Hyper	20	.0	.012	.014	.042	.049	.125	.061	.109	.016
		.4	.015	.015	.047	.057	.129	.062	.105	.013
		.8	.027	.015	.054	.084	.138	.069	.115	.011
	60	.0	.007	.012	.020	.032	.090	.035	.141	.022
		.4	.005	.013	.022	.035	.091	.035	.140	.019
		.8	.009	.016	.030	.043	.110	.057	.150	.018

Anmerkung: Vorpalte 1 gibt die Verteilung an, Vorpalte 2 den Stichprobenumfang N , Vorpalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Tabelle D-48. Mittlere Intervalllänge

			Perc ($\hat{d}_R^{S_2}$)		BCA ($\hat{d}_R^{S_2}$)		Perc-t ($\hat{d}_R^{S_2}$)		NC ($d_{s,R}^{S_2}$)	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	20	.0	2.114	2.150	2.143	2.197	1.765	1.873	1.681	1.701
		.4	1.874	1.994	1.917	2.048	1.564	1.762	1.452	1.569
		.8	1.533	1.802	1.589	1.855	1.289	1.583	1.123	1.386
	60	.0	.994	1.000	.993	1.001	.950	.968	.909	.916
		.4	.866	.923	.866	.923	.827	.896	.780	.838
		.8	.681	.822	.681	.822	.651	.797	.591	.738
Gleich	20	.0	2.438	2.405	2.398	2.364	1.983	2.076	1.991	2.012
		.4	2.070	2.191	2.039	2.156	1.678	1.913	1.681	1.834
		.8	1.590	1.966	1.567	1.930	1.261	1.675	1.238	1.597
	60	.0	1.133	1.128	1.126	1.120	1.082	1.097	1.085	1.094
		.4	.948	1.016	.940	1.007	.903	.990	.908	.984
		.8	.686	.890	.677	.882	.648	.860	.655	.856
g-h-1	20	.0	1.810	1.873	1.886	1.966	1.574	1.715	1.308	1.323
		.4	1.667	1.760	1.754	1.871	1.473	1.660	1.150	1.231
		.8	1.458	1.626	1.540	1.718	1.340	1.529	.943	1.111
	60	.0	.840	.848	.844	.856	.810	.829	.701	.707
		.4	.755	.792	.760	.798	.731	.777	.610	.650
		.8	.646	.732	.652	.739	.632	.722	.493	.589
g-h-2	20	.0	2.632	2.273	2.481	2.180	2.139	2.040	1.547	1.481
		.4	2.270	1.940	2.144	1.889	1.877	1.793	1.327	1.336
		.8	1.857	1.584	1.771	1.577	1.621	1.439	1.059	1.161
	60	.0	1.170	.996	1.149	.983	1.093	.967	.809	.772
		.4	1.012	.858	.993	.847	.948	.842	.692	.696
		.8	.815	.700	.800	.693	.775	.688	.542	.613
g-h-3	20	.0	1.890	1.815	1.924	1.883	1.667	1.703	1.274	1.249
		.4	1.713	1.677	1.772	1.768	1.539	1.619	1.115	1.158
		.8	1.488	1.489	1.551	1.571	1.409	1.407	.923	1.033
	60	.0	.872	.825	.873	.829	.845	.812	.676	.662
		.4	.778	.752	.780	.757	.757	.744	.588	.607
		.8	.663	.671	.665	.676	.652	.662	.480	.549
g-h-4	20	.0	3.351	2.838	3.116	2.680	2.441	2.382	1.873	1.830
		.4	2.794	2.346	2.609	2.249	2.076	2.020	1.585	1.637
		.8	2.177	1.833	2.035	1.777	1.660	1.569	1.204	1.394
	60	.0	1.373	1.174	1.344	1.154	1.251	1.123	.984	.951
		.4	1.151	.992	1.125	.976	1.051	.964	.824	.855
		.8	.874	.777	.852	.765	.802	.758	.613	.739
Hyper	20	.0	2.584	2.509	2.448	2.412	2.834	2.686	1.835	2.420
		.4	2.269	2.307	2.160	2.193	2.475	2.523	1.609	2.254
		.8	1.812	2.095	1.700	1.933	2.151	2.259	1.266	2.029
	60	.0	1.553	1.537	1.526	1.472	1.641	1.450	.997	1.288
		.4	1.367	1.413	1.344	1.337	1.483	1.339	.863	1.194
		.8	1.114	1.266	1.083	1.168	1.277	1.187	.674	1.070

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf fünf Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 2.2.4.3

Anhang E Ergebnisse der Studie 3

Tabelle E-1. $|CP - .95|$

			MEE-HAL-1		MEE-HAL-2		CLI-HAL-Z		CLI-HAL-T		CLI-LOG-Z		CLI-LOG-T	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	12 bis 24	1:3	.017	.047	.017	.062	.012	.080	.016	.067	.023	.031	.036	.027
		1:1	.011	.014	.010	.016	.015	.028	.025	.028	.030	.023	.040	.037
		3:1	.	.014	.	.012	.	.018	.	.027	.	.032	.	.041
	28 bis 40	1:3	.010	.042	.011	.052	.006	.049	.006	.044	.010	.032	.018	.027
		1:1	.005	.012	.005	.013	.006	.012	.013	.015	.016	.012	.023	.020
		3:1	.	.005	.	.005	.	.008	.	.016	.	.017	.	.025
Gleich	12 bis 24	1:3	.016	.042	.016	.062	.016	.095	.016	.072	.022	.036	.034	.027
		1:1	.013	.012	.012	.015	.016	.022	.028	.022	.032	.023	.040	.036
		3:1	.	.014	.	.013	.	.019	.	.028	.	.032	.	.041
	28 bis 40	1:3	.005	.028	.006	.037	.003	.034	.008	.031	.012	.024	.019	.022
		1:1	.004	.006	.004	.006	.008	.006	.016	.013	.017	.013	.024	.024
		3:1	.	.005	.	.004	.	.009	.	.018	.	.018	.	.026
g-h-1	12 bis 24	1:3	.020	.046	.021	.060	.020	.073	.019	.061	.025	.028	.037	.028
		1:1	.011	.015	.012	.019	.021	.032	.025	.029	.026	.023	.038	.036
		3:1	.	.012	.	.011	.	.019	.	.028	.	.031	.	.041
	28 bis 40	1:3	.015	.047	.016	.055	.011	.051	.008	.044	.007	.030	.017	.023
		1:1	.008	.013	.008	.015	.009	.014	.013	.018	.013	.013	.021	.019
		3:1	.	.005	.	.005	.	.009	.	.016	.	.017	.	.024
g-h-2	12 bis 24	1:3	.014	.034	.012	.040	.012	.037	.021	.022	.026	.019	.038	.029
		1:1	.013	.010	.013	.009	.027	.010	.030	.023	.027	.028	.038	.039
		3:1	.	.011	.	.012	.	.018	.	.027	.	.031	.	.040
	28 bis 40	1:3	.003	.013	.003	.014	.006	.009	.014	.007	.016	.010	.024	.018
		1:1	.009	.003	.009	.004	.010	.007	.015	.017	.014	.017	.021	.025
		3:1	.	.009	.	.010	.	.011	.	.016	.	.014	.	.022
g-h-3	12 bis 24	1:3	.018	.044	.017	.053	.014	.070	.018	.047	.026	.024	.038	.028
		1:1	.011	.014	.013	.016	.022	.028	.025	.029	.026	.025	.037	.036
		3:1	.	.012	.	.011	.	.021	.	.027	.	.031	.	.041
	28 bis 40	1:3	.009	.042	.009	.047	.004	.043	.006	.034	.010	.022	.018	.016
		1:1	.009	.014	.009	.014	.009	.014	.013	.017	.012	.013	.020	.020
		3:1	.	.005	.	.006	.	.009	.	.014	.	.016	.	.024
g-h-4	12 bis 24	1:3	.013	.033	.012	.036	.012	.037	.022	.024	.026	.019	.038	.027
		1:1	.011	.011	.012	.010	.021	.011	.028	.024	.027	.028	.038	.040
		3:1	.	.011	.	.014	.	.028	.	.029	.	.029	.	.039
	28 bis 40	1:3	.004	.007	.004	.007	.009	.005	.016	.009	.017	.013	.024	.021
		1:1	.006	.003	.006	.003	.008	.008	.014	.017	.014	.017	.022	.026
		3:1	.	.010	.	.011	.	.012	.	.016	.	.014	.	.021
Hyper	12 bis 24	1:3	.015	.040	.015	.055	.017	.096	.013	.077	.019	.044	.032	.031
		1:1	.015	.014	.013	.012	.021	.022	.025	.023	.023	.022	.035	.035
		3:1	.	.020	.	.019	.	.016	.	.026	.	.030	.	.040
	28 bis 40	1:3	.004	.021	.004	.030	.004	.036	.006	.033	.008	.027	.017	.025
		1:1	.006	.008	.008	.008	.009	.010	.013	.016	.011	.013	.018	.022
		3:1	.	.009	.	.009	.	.006	.	.015	.	.015	.	.023

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer (siehe Kapitel 3.1.4) aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf 20 Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 3.1.5

Tabelle E-2. Mittlere Intervalllänge

			MEE-HAL-1		MEE-HAL-2		CLI-HAL-Z		CLI-HAL-T		CLI-LOG-Z		CLI-LOG-T	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	12 bis 24	1:3	.474	.537	.472	.529	.487	.528	.526	.597	.541	.606	.594	.695
		1:1	.435	.455	.434	.454	.447	.461	.486	.513	.486	.506	.538	.574
		3:1	.	.425	.	.425	.	.443	.	.481	.	.483	.	.533
	28 bis 40	1:3	.357	.435	.357	.434	.364	.436	.379	.474	.382	.470	.399	.516
		1:1	.319	.343	.319	.343	.325	.346	.338	.367	.337	.361	.352	.386
		3:1	.	.310	.	.310	.	.316	.	.330	.	.327	.	.343
Gleich	12 bis 24	1:3	.475	.542	.473	.533	.488	.531	.527	.602	.543	.609	.595	.701
		1:1	.431	.455	.430	.454	.445	.462	.482	.514	.482	.506	.531	.575
		3:1	.	.423	.	.422	.	.442	.	.480	.	.481	.	.531
	28 bis 40	1:3	.356	.438	.356	.437	.363	.439	.377	.478	.380	.473	.397	.521
		1:1	.315	.342	.315	.342	.321	.346	.334	.367	.332	.361	.347	.385
		3:1	.	.307	.	.307	.	.314	.	.328	.	.325	.	.340
g-h-1	12 bis 24	1:3	.475	.530	.473	.523	.486	.524	.527	.588	.541	.599	.596	.683
		1:1	.440	.455	.439	.453	.450	.460	.491	.510	.491	.505	.545	.571
		3:1	.	.431	.	.430	.	.448	.	.486	.	.489	.	.539
	28 bis 40	1:3	.362	.429	.362	.428	.368	.430	.383	.463	.386	.462	.404	.503
		1:1	.326	.342	.326	.342	.331	.346	.345	.366	.344	.361	.360	.384
		3:1	.	.316	.	.316	.	.323	.	.337	.	.335	.	.351
g-h-2	12 bis 24	1:3	.448	.507	.447	.503	.465	.514	.503	.568	.510	.578	.560	.648
		1:1	.442	.441	.441	.441	.451	.452	.494	.497	.493	.493	.550	.551
		3:1	.	.446	.	.445	.	.462	.	.500	.	.509	.	.560
	28 bis 40	1:3	.330	.395	.330	.395	.337	.400	.351	.424	.350	.423	.366	.452
		1:1	.328	.326	.328	.326	.333	.332	.349	.348	.346	.344	.365	.362
		3:1	.	.329	.	.329	.	.336	.	.350	.	.350	.	.367
g-h-3	12 bis 24	1:3	.467	.524	.465	.518	.480	.522	.520	.583	.532	.594	.585	.674
		1:1	.440	.452	.439	.451	.450	.458	.491	.507	.491	.503	.545	.567
		3:1	.	.435	.	.434	.	.451	.	.489	.	.494	.	.544
	28 bis 40	1:3	.353	.420	.353	.419	.359	.422	.374	.452	.376	.452	.393	.488
		1:1	.326	.339	.326	.339	.331	.343	.345	.362	.344	.358	.360	.379
		3:1	.	.320	.	.320	.	.326	.	.340	.	.339	.	.355
g-h-4	12 bis 24	1:3	.448	.508	.447	.504	.466	.515	.503	.569	.511	.579	.560	.650
		1:1	.442	.440	.441	.439	.451	.452	.495	.496	.492	.492	.549	.549
		3:1	.	.448	.	.446	.	.463	.	.501	.	.511	.	.562
	28 bis 40	1:3	.330	.394	.330	.394	.337	.399	.351	.424	.350	.422	.366	.452
		1:1	.327	.326	.327	.326	.332	.331	.348	.347	.345	.343	.363	.361
		3:1	.	.330	.	.330	.	.337	.	.352	.	.353	.	.369
Hyper	12 bis 24	1:3	.470	.553	.466	.541	.466	.524	.505	.603	.511	.598	.562	.698
		1:1	.442	.462	.439	.459	.436	.453	.477	.507	.473	.493	.525	.563
		3:1	.	.427	.	.426	.	.425	.	.463	.	.458	.	.506
	28 bis 40	1:3	.348	.441	.348	.439	.348	.436	.362	.478	.363	.468	.379	.520
		1:1	.322	.344	.322	.344	.323	.342	.337	.364	.334	.356	.350	.380
		3:1	.	.309	.	.309	.	.308	.	.322	.	.318	.	.333

Anmerkung: Vorpalte 1 gibt die Verteilung an, Vorpalte 2 den Stichprobenumfang N , Vorpalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer (siehe Kapitel 3.1.4) aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf 20 Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 3.1.5

Tabelle E-3. $|CP - .95|$

			VD-WALD		VD-HAL-Z		VD-HAL-T		VD-LOG-Z		VD-LOG-Z	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	12 bis 24	1:3	.023	.084	.015	.017	.032	.038	.026	.034	.039	.046
		1:1	.016	.022	.025	.014	.034	.028	.033	.025	.040	.036
		3:1	.	.021	.	.007	.	.019	.	.014	.	.026
	28 bis 40	1:3	.009	.040	.015	.011	.026	.028	.021	.023	.032	.037
		1:1	.014	.013	.020	.010	.025	.015	.025	.013	.030	.021
		3:1	.	.017	.	.012	.	.011	.	.011	.	.013
Gleich	12 bis 24	1:3	.022	.078	.011	.019	.026	.038	.020	.035	.034	.045
		1:1	.020	.024	.025	.016	.034	.029	.033	.026	.039	.036
		3:1	.	.022	.	.008	.	.017	.	.012	.	.025
	28 bis 40	1:3	.010	.030	.013	.012	.024	.028	.019	.024	.029	.037
		1:1	.017	.019	.020	.015	.025	.016	.025	.015	.030	.021
		3:1	.	.024	.	.016	.	.014	.	.015	.	.014
g-h-1	12 bis 24	1:3	.024	.069	.016	.016	.035	.038	.029	.034	.042	.046
		1:1	.013	.021	.023	.016	.032	.028	.032	.026	.039	.036
		3:1	.	.019	.	.007	.	.021	.	.015	.	.029
	28 bis 40	1:3	.009	.042	.014	.012	.026	.027	.021	.023	.033	.037
		1:1	.010	.009	.018	.009	.023	.016	.023	.015	.028	.022
		3:1	.	.014	.	.011	.	.012	.	.011	.	.016
g-h-2	12 bis 24	1:3	.020	.044	.007	.017	.022	.034	.017	.031	.029	.042
		1:1	.013	.021	.021	.022	.031	.032	.029	.031	.038	.039
		3:1	.	.019	.	.013	.	.027	.	.021	.	.035
	28 bis 40	1:3	.012	.017	.010	.015	.019	.027	.015	.024	.024	.036
		1:1	.008	.019	.016	.017	.022	.021	.021	.020	.027	.026
		3:1	.	.015	.	.018	.	.021	.	.018	.	.024
g-h-3	12 bis 24	1:3	.023	.063	.016	.017	.033	.038	.027	.035	.040	.046
		1:1	.013	.020	.023	.016	.032	.029	.031	.028	.038	.037
		3:1	.	.020	.	.011	.	.025	.	.019	.	.032
	28 bis 40	1:3	.010	.032	.014	.013	.025	.028	.021	.023	.031	.037
		1:1	.009	.009	.017	.011	.023	.018	.023	.016	.028	.024
		3:1	.	.013	.	.012	.	.015	.	.013	.	.020
g-h-4	12 bis 24	1:3	.019	.042	.007	.016	.020	.031	.015	.028	.028	.038
		1:1	.013	.023	.022	.020	.031	.031	.030	.030	.038	.038
		3:1	.	.020	.	.012	.	.022	.	.017	.	.030
	28 bis 40	1:3	.013	.017	.010	.013	.019	.026	.014	.024	.023	.035
		1:1	.010	.022	.016	.019	.022	.021	.021	.020	.027	.025
		3:1	.	.014	.	.017	.	.020	.	.018	.	.022
Hyper	12 bis 24	1:3	.018	.081	.013	.021	.029	.036	.023	.029	.036	.042
		1:1	.009	.029	.022	.020	.032	.025	.030	.022	.039	.034
		3:1	.	.038	.	.029	.	.025	.	.026	.	.027
	28 bis 40	1:3	.008	.028	.013	.014	.024	.029	.020	.026	.029	.038
		1:1	.009	.023	.018	.021	.024	.021	.024	.021	.029	.022
		3:1	.	.042	.	.038	.	.036	.	.037	.	.035

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer (siehe Kapitel 3.1.4) aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf 20 Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 3.1.5

Tabelle E-4. Mittlere Intervalllänge

			VD-WALD		VD-HAL-Z		VD-HAL-T		VD-LOG-Z		VD-LOG-Z	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	12 bis 24	1:3	.589	.659	.515	.603	.592	.698	.590	.717	.682	.823
		1:1	.506	.506	.476	.482	.511	.522	.534	.545	.576	.591
		3:1	.	.463	.	.428	.	.472	.	.477	.	.529
	28 bis 40	1:3	.424	.510	.384	.464	.414	.510	.412	.519	.448	.574
		1:1	.363	.363	.346	.347	.358	.362	.366	.371	.381	.388
		3:1	.	.322	.	.306	.	.321	.	.322	.	.339
Gleich	12 bis 24	1:3	.591	.661	.517	.606	.596	.701	.592	.717	.686	.825
		1:1	.506	.507	.474	.480	.509	.520	.530	.541	.572	.588
		3:1	.	.456	.	.425	.	.468	.	.473	.	.525
	28 bis 40	1:3	.423	.513	.384	.464	.415	.510	.413	.515	.449	.570
		1:1	.363	.364	.345	.346	.357	.361	.364	.368	.379	.386
		3:1	.	.317	.	.303	.	.317	.	.319	.	.336
g-h-1	12 bis 24	1:3	.587	.653	.515	.594	.590	.686	.590	.705	.679	.809
		1:1	.506	.506	.480	.484	.515	.523	.539	.546	.581	.592
		3:1	.	.477	.	.436	.	.484	.	.486	.	.542
	28 bis 40	1:3	.425	.503	.384	.457	.415	.501	.414	.511	.450	.564
		1:1	.363	.363	.347	.348	.360	.363	.369	.372	.384	.389
		3:1	.	.333	.	.314	.	.331	.	.332	.	.350
g-h-2	12 bis 24	1:3	.538	.631	.472	.553	.534	.641	.529	.642	.604	.745
		1:1	.506	.507	.480	.475	.517	.513	.539	.534	.582	.578
		3:1	.	.517	.	.472	.	.533	.	.537	.	.609
	28 bis 40	1:3	.381	.479	.346	.418	.368	.456	.365	.451	.391	.497
		1:1	.363	.364	.347	.345	.361	.359	.370	.365	.385	.381
		3:1	.	.362	.	.345	.	.367	.	.370	.	.397
g-h-3	12 bis 24	1:3	.574	.648	.502	.582	.573	.674	.570	.688	.655	.792
		1:1	.505	.506	.480	.483	.515	.521	.539	.545	.581	.590
		3:1	.	.490	.	.446	.	.496	.	.499	.	.558
	28 bis 40	1:3	.414	.496	.373	.446	.401	.488	.399	.495	.432	.546
		1:1	.362	.363	.347	.348	.360	.362	.369	.371	.384	.388
		3:1	.	.342	.	.322	.	.340	.	.341	.	.362
g-h-4	12 bis 24	1:3	.538	.630	.473	.554	.536	.642	.531	.643	.607	.747
		1:1	.506	.507	.479	.474	.515	.511	.536	.531	.579	.577
		3:1	.	.515	.	.475	.	.536	.	.540	.	.613
	28 bis 40	1:3	.381	.479	.347	.417	.369	.455	.366	.450	.392	.496
		1:1	.363	.364	.347	.344	.360	.358	.368	.364	.384	.380
		3:1	.	.361	.	.346	.	.369	.	.372	.	.399
Hyper	12 bis 24	1:3	.566	.657	.486	.592	.559	.692	.550	.709	.637	.821
		1:1	.497	.496	.468	.464	.503	.506	.525	.522	.567	.571
		3:1	.	.431	.	.402	.	.444	.	.446	.	.495
	28 bis 40	1:3	.406	.513	.362	.454	.390	.500	.383	.503	.416	.560
		1:1	.358	.357	.340	.337	.353	.352	.360	.356	.375	.374
		3:1	.	.304	.	.290	.	.304	.	.304	.	.319

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer (siehe Kapitel 3.1.4) aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf 20 Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 3.1.5

Tabelle E-5. $|CP - .95|$

			BM-WALD		BM-HAL-Z		BM-HAL-T		BM-LOG-Z		BM-LOG-Z	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	12 bis 24	1:3	.072	.161	.035	.125	.016	.079	.011	.068	.028	.042
		1:1	.041	.064	.017	.042	.018	.031	.019	.026	.031	.028
		3:1	.	.033	.	.018	.	.023	.	.021	.	.034
	28 bis 40	1:3	.034	.068	.012	.055	.008	.045	.003	.042	.016	.035
		1:1	.020	.026	.006	.014	.010	.016	.010	.013	.018	.017
		3:1	.	.017	.	.006	.	.012	.	.011	.	.019
Gleich	12 bis 24	1:3	.063	.144	.037	.121	.016	.081	.011	.072	.025	.047
		1:1	.031	.053	.013	.033	.019	.026	.021	.021	.034	.025
		3:1	.	.027	.	.017	.	.023	.	.021	.	.034
	28 bis 40	1:3	.022	.051	.007	.038	.009	.030	.005	.029	.017	.027
		1:1	.015	.020	.004	.007	.011	.012	.012	.009	.020	.020
		3:1	.	.017	.	.004	.	.012	.	.012	.	.021
g-h-1	12 bis 24	1:3	.080	.163	.043	.118	.019	.074	.013	.061	.030	.036
		1:1	.054	.068	.030	.046	.025	.035	.021	.029	.028	.029
		3:1	.	.038	.	.022	.	.025	.	.021	.	.034
	28 bis 40	1:3	.043	.077	.018	.058	.010	.047	.006	.043	.013	.034
		1:1	.023	.025	.009	.015	.012	.018	.010	.015	.016	.019
		3:1	.	.019	.	.006	.	.012	.	.011	.	.019
g-h-2	12 bis 24	1:3	.041	.088	.018	.059	.015	.025	.014	.025	.029	.020
		1:1	.052	.041	.034	.014	.029	.016	.027	.017	.031	.031
		3:1	.	.054	.	.029	.	.025	.	.020	.	.034
	28 bis 40	1:3	.017	.031	.003	.015	.011	.008	.010	.007	.020	.018
		1:1	.021	.017	.010	.004	.014	.013	.011	.012	.017	.021
		3:1	.	.026	.	.011	.	.013	.	.011	.	.017
g-h-3	12 bis 24	1:3	.067	.146	.032	.102	.016	.058	.012	.048	.030	.026
		1:1	.054	.064	.030	.041	.025	.032	.022	.026	.028	.028
		3:1	.	.043	.	.022	.	.025	.	.020	.	.033
	28 bis 40	1:3	.035	.074	.011	.050	.007	.038	.003	.033	.016	.025
		1:1	.024	.025	.010	.015	.012	.017	.010	.014	.015	.018
		3:1	.	.022	.	.007	.	.012	.	.010	.	.018
g-h-4	12 bis 24	1:3	.036	.077	.015	.056	.016	.025	.013	.026	.030	.020
		1:1	.042	.035	.028	.012	.025	.017	.024	.018	.030	.033
		3:1	.	.054	.	.036	.	.030	.	.023	.	.033
	28 bis 40	1:3	.016	.022	.003	.008	.012	.011	.011	.008	.021	.022
		1:1	.020	.017	.006	.003	.011	.013	.010	.012	.018	.022
		3:1	.	.020	.	.013	.	.014	.	.012	.	.019
Hyper	12 bis 24	1:3	.059	.141	.030	.113	.013	.083	.010	.078	.025	.048
		1:1	.051	.048	.032	.035	.025	.026	.022	.022	.028	.026
		3:1	.	.032	.	.014	.	.022	.	.018	.	.034
	28 bis 40	1:3	.026	.051	.008	.040	.007	.033	.003	.033	.016	.027
		1:1	.023	.019	.009	.011	.012	.016	.010	.013	.015	.020
		3:1	.	.017	.	.005	.	.012	.	.011	.	.020

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer (siehe Kapitel 3.1.4) aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf 20 Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 3.1.5

Tabelle E-6. Mittlere Intervalllänge

			BM-WALD		BM-HAL-Z		BM-HAL-T		BM-LOG-Z		BM-LOG-Z	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	12 bis 24	1:3	.573	.651	.464	.512	.540	.617	.508	.573	.605	.707
		1:1	.498	.531	.431	.448	.469	.499	.463	.486	.512	.552
		3:1	.	.483	.	.420	.	.459	.	.451	.	.503
	28 bis 40	1:3	.413	.526	.357	.432	.388	.488	.373	.464	.409	.533
		1:1	.348	.381	.319	.342	.332	.363	.330	.357	.345	.381
		3:1	.	.336	.	.310	.	.322	.	.320	.	.334
Gleich	12 bis 24	1:3	.575	.655	.466	.515	.545	.624	.509	.578	.612	.715
		1:1	.493	.534	.427	.449	.464	.501	.459	.487	.506	.553
		3:1	.	.480	.	.418	.	.458	.	.449	.	.501
	28 bis 40	1:3	.410	.530	.355	.435	.387	.492	.372	.467	.409	.539
		1:1	.344	.383	.315	.342	.328	.363	.325	.356	.339	.380
		3:1	.	.333	.	.307	.	.320	.	.317	.	.331
g-h-1	12 bis 24	1:3	.574	.644	.465	.507	.539	.608	.508	.566	.603	.694
		1:1	.503	.528	.435	.447	.474	.496	.469	.485	.520	.549
		3:1	.	.493	.	.425	.	.468	.	.458	.	.513
	28 bis 40	1:3	.418	.517	.361	.426	.393	.480	.379	.456	.415	.522
		1:1	.354	.378	.325	.342	.339	.362	.338	.356	.353	.379
		3:1	.	.344	.	.316	.	.330	.	.327	.	.343
g-h-2	12 bis 24	1:3	.529	.618	.442	.494	.498	.588	.478	.547	.549	.665
		1:1	.507	.514	.436	.437	.479	.481	.471	.471	.526	.527
		3:1	.	.519	.	.438	.	.496	.	.475	.	.551
	28 bis 40	1:3	.372	.482	.330	.395	.350	.438	.342	.416	.364	.468
		1:1	.359	.363	.328	.326	.344	.343	.340	.338	.359	.356
		3:1	.	.361	.	.328	.	.350	.	.342	.	.367
g-h-3	12 bis 24	1:3	.562	.639	.459	.504	.527	.603	.500	.562	.587	.687
		1:1	.502	.524	.434	.445	.475	.492	.469	.482	.521	.543
		3:1	.	.500	.	.429	.	.475	.	.462	.	.522
	28 bis 40	1:3	.406	.508	.352	.418	.380	.469	.368	.446	.400	.508
		1:1	.354	.374	.326	.339	.340	.357	.338	.353	.354	.374
		3:1	.	.349	.	.320	.	.335	.	.331	.	.349
g-h-4	12 bis 24	1:3	.530	.618	.443	.495	.500	.591	.479	.548	.551	.669
		1:1	.507	.514	.436	.436	.479	.480	.471	.470	.525	.525
		3:1	.	.520	.	.439	.	.501	.	.477	.	.559
	28 bis 40	1:3	.372	.480	.330	.394	.350	.437	.342	.415	.364	.467
		1:1	.358	.362	.327	.325	.342	.342	.339	.337	.357	.355
		3:1	.	.362	.	.330	.	.352	.	.344	.	.371
Hyper	12 bis 24	1:3	.552	.653	.448	.511	.517	.622	.485	.572	.572	.711
		1:1	.491	.532	.424	.444	.463	.497	.455	.479	.505	.547
		3:1	.	.472	.	.407	.	.446	.	.434	.	.484
	28 bis 40	1:3	.395	.532	.343	.434	.370	.491	.357	.465	.387	.536
		1:1	.347	.384	.318	.340	.332	.361	.330	.353	.346	.377
		3:1	.	.331	.	.303	.	.317	.	.313	.	.328

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer (siehe Kapitel 3.1.4) aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf 20 Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 3.1.5

Tabelle E-7. $|\overline{CP} - .95|$

			NEW-M-3		NEW-M-5		ZHOU	
			1:1	3:1	1:1	3:1	1:1	3:1
Normal	12 bis 24	1:3	.018	.042	.008	.059	.084	.178
		1:1	.009	.012	.009	.012	.046	.066
		3:1	.	.025	.	.034	.	.032
	28 bis 40	1:3	.017	.045	.007	.068	.032	.065
		1:1	.006	.017	.006	.017	.014	.019
		3:1	.	.022	.	.033	.	.008
Gleich	12 bis 24	1:3	.018	.043	.010	.062	.079	.163
		1:1	.010	.013	.010	.013	.036	.054
		3:1	.	.025	.	.034	.	.029
	28 bis 40	1:3	.018	.049	.009	.070	.022	.045
		1:1	.008	.018	.008	.018	.012	.013
		3:1	.	.023	.	.033	.	.011
g-h-1	12 bis 24	1:3	.013	.033	.006	.049	.092	.176
		1:1	.005	.009	.005	.009	.060	.070
		3:1	.	.021	.	.031	.	.037
	28 bis 40	1:3	.013	.040	.004	.061	.039	.071
		1:1	.003	.016	.003	.016	.018	.020
		3:1	.	.018	.	.030	.	.009
g-h-2	12 bis 24	1:3	.028	.034	.023	.037	.049	.106
		1:1	.005	.013	.005	.013	.055	.043
		3:1	.	.021	.	.024	.	.058
	28 bis 40	1:3	.029	.037	.023	.039	.012	.029
		1:1	.004	.012	.004	.012	.014	.013
		3:1	.	.021	.	.024	.	.024
g-h-3	12 bis 24	1:3	.019	.028	.010	.042	.078	.160
		1:1	.006	.008	.006	.008	.060	.068
		3:1	.	.020	.	.030	.	.041
	28 bis 40	1:3	.018	.033	.008	.053	.032	.067
		1:1	.003	.011	.003	.011	.019	.022
		3:1	.	.018	.	.028	.	.013
g-h-4	12 bis 24	1:3	.029	.039	.022	.043	.050	.095
		1:1	.006	.013	.006	.013	.045	.038
		3:1	.	.019	.	.024	.	.060
	28 bis 40	1:3	.030	.041	.024	.042	.013	.019
		1:1	.003	.014	.003	.014	.009	.015
		3:1	.	.024	.	.023	.	.022
Hyper	12 bis 24	1:3	.024	.048	.017	.064	.073	.152
		1:1	.008	.019	.008	.019	.054	.047
		3:1	.	.027	.	.034	.	.027
	28 bis 40	1:3	.024	.052	.015	.066	.023	.047
		1:1	.005	.018	.005	.018	.016	.016
		3:1	.	.025	.	.034	.	.008

Anmerkung: Vorpalte 1 gibt die Verteilung an, Vorpalte 2 den Stichprobenumfang N , Vorpalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer (siehe Kapitel 3.1.4) aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf 20 Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 3.1.5

Tabelle E-8. Mittlere Intervalllänge

			NEW-M-3		NEW-M-5		ZHOU	
			1:1	3:1	1:1	3:1	1:1	3:1
Normal	12 bis 24	1:3	.473	.460	.467	.453	.504	.583
		1:1	.420	.417	.420	.417	.465	.495
		3:1	.	.464	.	.473	.	.451
	28 bis 40	1:3	.367	.360	.356	.349	.381	.481
		1:1	.315	.313	.315	.313	.335	.362
		3:1	.	.346	.	.359	.	.323
Gleich	12 bis 24	1:3	.474	.460	.468	.453	.496	.584
		1:1	.421	.417	.421	.417	.456	.494
		3:1	.	.465	.	.474	.	.445
	28 bis 40	1:3	.367	.360	.357	.349	.377	.484
		1:1	.315	.314	.315	.314	.330	.362
		3:1	.	.346	.	.359	.	.320
g-h-1	12 bis 24	1:3	.472	.461	.466	.454	.511	.579
		1:1	.419	.417	.419	.417	.473	.494
		3:1	.	.464	.	.473	.	.459
	28 bis 40	1:3	.367	.360	.356	.349	.387	.473
		1:1	.315	.313	.315	.313	.342	.361
		3:1	.	.345	.	.359	.	.331
g-h-2	12 bis 24	1:3	.477	.467	.471	.461	.476	.547
		1:1	.419	.419	.419	.419	.473	.474
		3:1	.	.462	.	.471	.	.465
	28 bis 40	1:3	.369	.364	.358	.353	.349	.434
		1:1	.314	.315	.314	.315	.344	.344
		3:1	.	.344	.	.358	.	.343
g-h-3	12 bis 24	1:3	.474	.463	.468	.456	.502	.574
		1:1	.419	.417	.419	.417	.473	.490
		3:1	.	.463	.	.472	.	.464
	28 bis 40	1:3	.367	.361	.357	.350	.377	.463
		1:1	.314	.314	.314	.314	.342	.357
		3:1	.	.345	.	.358	.	.335
g-h-4	12 bis 24	1:3	.477	.467	.471	.461	.474	.542
		1:1	.419	.419	.419	.419	.472	.470
		3:1	.	.462	.	.471	.	.460
	28 bis 40	1:3	.369	.364	.358	.354	.349	.432
		1:1	.315	.315	.315	.315	.343	.343
		3:1	.	.344	.	.358	.	.344
Hyper	12 bis 24	1:3	.475	.459	.469	.453	.489	.587
		1:1	.420	.417	.420	.417	.460	.493
		3:1	.	.465	.	.474	.	.438
	28 bis 40	1:3	.368	.360	.357	.349	.366	.486
		1:1	.315	.314	.315	.314	.334	.361
		3:1	.	.346	.	.359	.	.317

Anmerkung: Vorpalte 1 gibt die Verteilung an, Vorpalte 2 den Stichprobenumfang N , Vorpalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer (siehe Kapitel 3.1.4) aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf 20 Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 3.1.5

Tabelle E-9. $|CP - .95|$

			S-BM-Z-1		S-BM-T-1		S-BM-Z-2		S-BM-T-2		S-BM-Z-3		S-BM-T-3	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	12 bis 24	1:3	.005	.016	.022	.024	.013	.006	.031	.031	.006	.008	.027	.028
		1:1	.011	.006	.025	.022	.021	.014	.032	.029	.016	.009	.029	.027
		3:1	.	.010	.	.025	.	.024	.	.035	.	.018	.	.031
	28 bis 40	1:3	.002	.011	.015	.017	.006	.006	.019	.020	.003	.008	.017	.019
		1:1	.006	.003	.015	.012	.012	.005	.018	.016	.009	.003	.017	.014
		3:1	.	.007	.	.015	.	.013	.	.020	.	.010	.	.018
Gleich	12 bis 24	1:3	.009	.028	.016	.015	.009	.011	.026	.021	.005	.020	.022	.018
		1:1	.012	.005	.024	.017	.021	.011	.031	.026	.017	.006	.029	.022
		3:1	.	.012	.	.028	.	.026	.	.037	.	.020	.	.033
	28 bis 40	1:3	.005	.021	.010	.007	.003	.015	.015	.010	.003	.017	.013	.008
		1:1	.009	.004	.016	.009	.013	.004	.020	.013	.011	.003	.018	.011
		3:1	.	.008	.	.016	.	.014	.	.022	.	.011	.	.019
g-h-1	12 bis 24	1:3	.007	.011	.026	.028	.017	.010	.035	.035	.010	.009	.031	.032
		1:1	.011	.007	.026	.025	.021	.018	.033	.032	.016	.012	.031	.029
		3:1	.	.010	.	.025	.	.024	.	.035	.	.017	.	.030
	28 bis 40	1:3	.003	.005	.018	.022	.007	.003	.022	.025	.005	.004	.020	.024
		1:1	.006	.003	.015	.014	.010	.007	.019	.018	.008	.005	.017	.016
		3:1	.	.007	.	.016	.	.013	.	.021	.	.010	.	.019
g-h-2	12 bis 24	1:3	.006	.020	.019	.007	.018	.010	.029	.016	.012	.011	.025	.012
		1:1	.014	.009	.029	.025	.023	.019	.036	.032	.018	.014	.033	.028
		3:1	.	.017	.	.035	.	.027	.	.040	.	.021	.	.038
	28 bis 40	1:3	.008	.010	.017	.007	.013	.007	.021	.011	.011	.008	.019	.009
		1:1	.008	.007	.017	.016	.011	.011	.021	.020	.010	.009	.019	.018
		3:1	.	.011	.	.023	.	.016	.	.027	.	.013	.	.025
g-h-3	12 bis 24	1:3	.005	.011	.023	.025	.017	.009	.032	.033	.010	.007	.029	.030
		1:1	.012	.008	.026	.025	.022	.018	.034	.032	.017	.012	.031	.029
		3:1	.	.011	.	.027	.	.025	.	.036	.	.018	.	.033
	28 bis 40	1:3	.003	.003	.017	.022	.009	.003	.021	.025	.006	.003	.020	.023
		1:1	.006	.004	.015	.016	.010	.009	.019	.019	.008	.006	.017	.017
		3:1	.	.007	.	.016	.	.012	.	.021	.	.010	.	.019
g-h-4	12 bis 24	1:3	.008	.020	.019	.007	.018	.015	.028	.015	.012	.015	.025	.011
		1:1	.013	.009	.028	.023	.022	.018	.034	.031	.017	.013	.032	.028
		3:1	.	.015	.	.033	.	.026	.	.039	.	.019	.	.037
	28 bis 40	1:3	.010	.014	.017	.006	.014	.012	.021	.009	.012	.012	.019	.007
		1:1	.007	.009	.016	.016	.011	.012	.019	.019	.009	.010	.018	.017
		3:1	.	.012	.	.021	.	.014	.	.025	.	.013	.	.023
Hyper	12 bis 24	1:3	.088	.207	.038	.146	.035	.026	.042	.029	.009	.046	.027	.024
		1:1	.081	.150	.043	.116	.039	.041	.043	.045	.017	.009	.029	.022
		3:1	.	.103	.	.041	.	.047	.	.049	.	.032	.	.044
	28 bis 40	1:3	.107	.201	.073	.167	.036	.026	.040	.031	.005	.034	.016	.017
		1:1	.120	.169	.093	.152	.039	.040	.042	.042	.011	.007	.019	.012
		3:1	.	.139	.	.105	.	.045	.	.047	.	.027	.	.035

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer (siehe Kapitel 3.1.4) aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf 20 Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 3.1.5

Tabelle E-10. Mittlere Intervalllänge

			S-BM-Z-1		S-BM-T-1		S-BM-Z-2		S-BM-T-2		S-BM-Z-3		S-BM-T-3	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	12 bis 24	1:3	.460	.527	.537	.637	.498	.563	.572	.671	.479	.544	.555	.654
		1:1	.425	.440	.463	.490	.455	.469	.492	.519	.439	.454	.477	.504
		3:1	.	.408	.	.450	.	.447	.	.488	.	.428	.	.469
	28 bis 40	1:3	.356	.429	.387	.486	.366	.438	.397	.495	.361	.434	.392	.491
		1:1	.318	.338	.331	.358	.325	.345	.338	.365	.322	.342	.335	.362
		3:1	.	.306	.	.319	.	.316	.	.329	.	.311	.	.324
Gleich	12 bis 24	1:3	.458	.525	.536	.631	.496	.561	.570	.665	.477	.542	.553	.649
		1:1	.423	.440	.459	.489	.453	.469	.489	.518	.437	.454	.474	.504
		3:1	.	.405	.	.449	.	.444	.	.486	.	.425	.	.467
	28 bis 40	1:3	.354	.430	.386	.486	.364	.439	.395	.495	.359	.435	.390	.491
		1:1	.315	.338	.328	.359	.322	.346	.335	.366	.319	.342	.331	.363
		3:1	.	.303	.	.316	.	.313	.	.325	.	.308	.	.320
g-h-1	12 bis 24	1:3	.463	.523	.541	.632	.502	.560	.576	.667	.483	.541	.559	.650
		1:1	.429	.440	.468	.489	.458	.469	.498	.518	.443	.454	.483	.503
		3:1	.	.414	.	.459	.	.453	.	.497	.	.434	.	.478
	28 bis 40	1:3	.360	.424	.392	.480	.370	.433	.402	.489	.365	.429	.397	.485
		1:1	.323	.337	.336	.356	.330	.344	.344	.363	.326	.341	.340	.360
		3:1	.	.312	.	.327	.	.322	.	.336	.	.317	.	.332
g-h-2	12 bis 24	1:3	.473	.525	.569	.635	.512	.563	.603	.671	.493	.544	.586	.653
		1:1	.430	.439	.472	.490	.459	.468	.501	.519	.444	.453	.486	.504
		3:1	.	.416	.	.465	.	.455	.	.503	.	.436	.	.484
	28 bis 40	1:3	.349	.416	.383	.471	.359	.425	.393	.481	.354	.421	.388	.476
		1:1	.324	.331	.339	.350	.331	.338	.346	.357	.327	.335	.343	.353
		3:1	.	.319	.	.337	.	.329	.	.347	.	.324	.	.342
g-h-3	12 bis 24	1:3	.466	.525	.547	.636	.504	.562	.583	.671	.485	.543	.565	.654
		1:1	.429	.440	.469	.489	.458	.469	.498	.518	.443	.454	.483	.504
		3:1	.	.414	.	.459	.	.453	.	.496	.	.434	.	.478
	28 bis 40	1:3	.357	.421	.389	.477	.366	.431	.398	.487	.361	.426	.394	.482
		1:1	.323	.336	.337	.355	.330	.343	.344	.362	.327	.339	.340	.358
		3:1	.	.314	.	.329	.	.324	.	.339	.	.319	.	.334
g-h-4	12 bis 24	1:3	.473	.526	.570	.634	.511	.564	.603	.671	.492	.545	.586	.653
		1:1	.429	.440	.470	.491	.458	.469	.500	.520	.443	.454	.485	.505
		3:1	.	.415	.	.465	.	.454	.	.503	.	.435	.	.484
	28 bis 40	1:3	.349	.418	.383	.474	.358	.428	.392	.484	.353	.423	.387	.479
		1:1	.323	.331	.338	.350	.330	.338	.345	.357	.326	.335	.342	.354
		3:1	.	.318	.	.336	.	.328	.	.346	.	.323	.	.341
Hyper	12 bis 24	1:3	.384	.408	.482	.503	.603	.644	.688	.736	.492	.524	.583	.621
		1:1	.338	.342	.377	.385	.545	.566	.581	.609	.440	.451	.479	.497
		3:1	.	.333	.	.402	.	.572	.	.634	.	.444	.	.508
	28 bis 40	1:3	.274	.325	.309	.372	.465	.521	.498	.570	.366	.422	.401	.471
		1:1	.236	.250	.250	.267	.424	.436	.437	.454	.325	.340	.338	.358
		3:1	.	.230	.	.249	.	.424	.	.442	.	.323	.	.342

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer (siehe Kapitel 3.1.4) aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf 20 Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 3.1.5

Tabelle E-11. $|\overline{CP} - .95|$

			S-CLI-Z-1		S-CLI-T-1		S-CLI-Z-2		S-CLI-T-2		S-CLI-Z-3		S-CLI-T-3	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	12 bis 24	1:3	.008	.009	.022	.015	.021	.009	.031	.022	.015	.006	.027	.018
		1:1	.017	.010	.030	.027	.026	.020	.035	.033	.022	.015	.033	.030
		3:1	.	.020	.	.032	.	.031	.	.039	.	.026	.	.036
	28 bis 40	1:3	.005	.007	.013	.008	.010	.004	.018	.011	.008	.005	.015	.009
		1:1	.011	.004	.018	.015	.015	.007	.022	.018	.013	.006	.020	.017
		3:1	.	.012	.	.020	.	.017	.	.024	.	.015	.	.022
Gleich	12 bis 24	1:3	.005	.019	.019	.006	.017	.008	.028	.013	.011	.011	.024	.008
		1:1	.017	.008	.029	.022	.026	.017	.035	.030	.023	.012	.032	.027
		3:1	.	.022	.	.034	.	.032	.	.040	.	.028	.	.037
	28 bis 40	1:3	.003	.016	.009	.006	.007	.011	.014	.003	.005	.013	.012	.004
		1:1	.012	.003	.019	.013	.016	.006	.023	.016	.014	.004	.021	.014
		3:1	.	.013	.	.021	.	.018	.	.026	.	.015	.	.023
g-h-1	12 bis 24	1:3	.012	.011	.026	.021	.024	.014	.034	.028	.018	.012	.030	.024
		1:1	.017	.013	.031	.029	.026	.022	.037	.035	.022	.018	.034	.032
		3:1	.	.018	.	.031	.	.030	.	.038	.	.025	.	.035
	28 bis 40	1:3	.007	.003	.016	.013	.012	.005	.020	.017	.009	.003	.018	.015
		1:1	.010	.006	.019	.017	.014	.010	.022	.020	.012	.008	.020	.019
		3:1	.	.012	.	.020	.	.017	.	.024	.	.015	.	.022
g-h-2	12 bis 24	1:3	.013	.011	.026	.008	.025	.009	.034	.017	.020	.008	.030	.013
		1:1	.019	.014	.033	.029	.027	.024	.038	.035	.023	.019	.036	.032
		3:1	.	.022	.	.034	.	.033	.	.040	.	.028	.	.038
	28 bis 40	1:3	.011	.008	.020	.007	.017	.007	.024	.010	.014	.007	.022	.008
		1:1	.011	.010	.021	.019	.014	.014	.024	.022	.013	.012	.023	.021
		3:1	.	.015	.	.023	.	.020	.	.027	.	.018	.	.026
g-h-3	12 bis 24	1:3	.011	.010	.025	.019	.024	.014	.033	.027	.019	.010	.030	.023
		1:1	.018	.014	.031	.029	.026	.023	.036	.035	.022	.018	.034	.032
		3:1	.	.020	.	.033	.	.031	.	.039	.	.026	.	.037
	28 bis 40	1:3	.007	.002	.016	.014	.013	.006	.020	.017	.011	.004	.018	.015
		1:1	.010	.007	.018	.018	.014	.012	.022	.022	.012	.010	.020	.020
		3:1	.	.011	.	.020	.	.017	.	.024	.	.014	.	.022
g-h-4	12 bis 24	1:3	.013	.014	.025	.010	.024	.014	.033	.017	.019	.014	.030	.014
		1:1	.018	.014	.032	.028	.027	.023	.037	.034	.023	.019	.035	.032
		3:1	.	.022	.	.033	.	.031	.	.039	.	.027	.	.037
	28 bis 40	1:3	.013	.013	.019	.009	.017	.012	.023	.011	.015	.012	.021	.010
		1:1	.010	.011	.019	.018	.014	.015	.022	.022	.012	.013	.021	.020
		3:1	.	.014	.	.021	.	.018	.	.025	.	.015	.	.023
Hyper	12 bis 24	1:3	.067	.185	.036	.154	.039	.026	.043	.030	.018	.039	.029	.026
		1:1	.065	.139	.032	.104	.041	.042	.045	.046	.022	.013	.033	.027
		3:1	.	.081	.	.052	.	.048	.	.049	.	.037	.	.043
	28 bis 40	1:3	.095	.194	.075	.180	.038	.026	.040	.029	.009	.030	.016	.022
		1:1	.108	.164	.082	.147	.040	.041	.043	.043	.014	.009	.022	.013
		3:1	.	.129	.	.109	.	.046	.	.047	.	.030	.	.035

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer (siehe Kapitel 3.1.4) aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf 20 Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 3.1.5

Tabelle E-12. Mittlere Intervalllänge

			S-CLI-Z-1		S-CLI-T-1		S-CLI-Z-2		S-CLI-T-2		S-CLI-Z-3		S-CLI-T-3	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	12 bis 24	1:3	.480	.540	.519	.600	.519	.576	.556	.631	.499	.557	.537	.614
		1:1	.440	.452	.478	.502	.469	.481	.508	.531	.454	.467	.493	.517
		3:1	.	.432	.	.470	.	.471	.	.509	.	.451	.	.489
	28 bis 40	1:3	.363	.433	.377	.466	.373	.442	.387	.475	.368	.438	.382	.471
		1:1	.324	.342	.337	.362	.331	.349	.344	.370	.327	.345	.341	.366
		3:1	.	.313	.	.327	.	.323	.	.337	.	.318	.	.332
Gleich	12 bis 24	1:3	.479	.537	.517	.594	.517	.574	.554	.627	.498	.555	.535	.609
		1:1	.438	.452	.475	.502	.468	.481	.505	.531	.453	.467	.490	.516
		3:1	.	.429	.	.468	.	.469	.	.507	.	.449	.	.487
	28 bis 40	1:3	.361	.434	.375	.467	.371	.443	.385	.477	.366	.439	.380	.472
		1:1	.321	.342	.334	.363	.328	.349	.341	.370	.325	.346	.338	.367
		3:1	.	.310	.	.324	.	.320	.	.334	.	.315	.	.329
g-h-1	12 bis 24	1:3	.483	.537	.524	.597	.521	.573	.561	.627	.502	.554	.542	.611
		1:1	.443	.452	.483	.501	.472	.481	.512	.530	.457	.466	.498	.516
		3:1	.	.437	.	.476	.	.476	.	.514	.	.456	.	.494
	28 bis 40	1:3	.366	.428	.382	.459	.376	.438	.391	.468	.371	.433	.387	.464
		1:1	.328	.341	.342	.360	.335	.348	.349	.368	.332	.345	.346	.364
		3:1	.	.319	.	.333	.	.329	.	.343	.	.324	.	.338
g-h-2	12 bis 24	1:3	.494	.540	.546	.606	.533	.578	.582	.640	.513	.558	.563	.622
		1:1	.443	.452	.486	.503	.473	.481	.516	.532	.458	.466	.501	.518
		3:1	.	.439	.	.478	.	.478	.	.516	.	.458	.	.496
	28 bis 40	1:3	.356	.421	.375	.454	.366	.430	.385	.464	.361	.425	.380	.459
		1:1	.329	.336	.344	.354	.336	.343	.352	.362	.333	.339	.348	.358
		3:1	.	.326	.	.340	.	.336	.	.350	.	.331	.	.345
g-h-3	12 bis 24	1:3	.486	.539	.529	.601	.524	.576	.565	.632	.505	.557	.547	.615
		1:1	.443	.452	.484	.502	.472	.481	.513	.531	.458	.467	.498	.516
		3:1	.	.437	.	.475	.	.476	.	.514	.	.456	.	.494
	28 bis 40	1:3	.363	.426	.379	.457	.373	.435	.389	.466	.368	.431	.384	.461
		1:1	.328	.340	.342	.359	.335	.347	.349	.366	.332	.343	.346	.363
		3:1	.	.321	.	.335	.	.331	.	.345	.	.326	.	.340
g-h-4	12 bis 24	1:3	.493	.541	.544	.607	.532	.579	.580	.641	.512	.559	.561	.623
		1:1	.443	.452	.485	.504	.472	.482	.514	.533	.458	.467	.500	.519
		3:1	.	.439	.	.477	.	.477	.	.515	.	.457	.	.495
	28 bis 40	1:3	.356	.423	.375	.459	.366	.433	.384	.468	.361	.428	.379	.463
		1:1	.328	.336	.343	.355	.335	.343	.351	.362	.332	.339	.347	.359
		3:1	.	.325	.	.339	.	.335	.	.349	.	.330	.	.344
Hyper	12 bis 24	1:3	.404	.423	.454	.477	.620	.658	.663	.707	.510	.537	.556	.591
		1:1	.352	.353	.393	.397	.558	.576	.594	.620	.454	.463	.494	.509
		3:1	.	.356	.	.394	.	.593	.	.631	.	.466	.	.506
	28 bis 40	1:3	.281	.328	.297	.357	.471	.525	.487	.553	.372	.425	.389	.455
		1:1	.242	.253	.256	.270	.429	.439	.442	.458	.330	.344	.343	.362
		3:1	.	.236	.	.250	.	.430	.	.444	.	.329	.	.344

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 das Verhältnis $n_1:n_2$; in der Kopfzeile sind die untersuchten Schätzer (siehe Kapitel 3.1.4) aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; jede Zelle basiert auf 20 Simulationsbedingungen, siehe hierzu das Untersuchungsdesign in Kapitel 3.1.5

Anhang F Ergebnisse der Studie 4

Tabelle F-1. $|\overline{CP} - .95|$

			MAX		Pratt		Blyth		MID-P		Wilson	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	6 bis 12	.0	.033	.033	.033	.033	.031	.032	.025	.026	.017	.016
		.4	.034	.035	.035	.036	.031	.033	.025	.026	.020	.018
		.8	.038	.034	.039	.035	.029	.030	.024	.026	.021	.019
	13 bis 20	.0	.027	.026	.027	.026	.028	.027	.017	.016	.012	.012
		.4	.030	.029	.030	.029	.027	.029	.020	.017	.014	.014
		.8	.033	.028	.033	.028	.033	.028	.027	.022	.022	.014
Gleich	6 bis 12	.0	.033	.034	.033	.034	.031	.032	.025	.024	.017	.016
		.4	.034	.036	.034	.036	.030	.033	.027	.025	.019	.018
		.8	.039	.036	.039	.036	.030	.034	.027	.030	.021	.028
	13 bis 20	.0	.027	.027	.027	.027	.028	.028	.017	.016	.014	.013
		.4	.028	.029	.028	.029	.027	.027	.018	.018	.014	.013
		.8	.034	.029	.034	.029	.032	.027	.029	.020	.023	.016
g-h-1	6 bis 12	.0	.033	.033	.033	.033	.032	.032	.025	.026	.017	.017
		.4	.035	.035	.035	.035	.032	.033	.026	.026	.020	.020
		.8	.037	.034	.037	.035	.033	.030	.029	.027	.030	.019
	13 bis 20	.0	.027	.026	.027	.026	.028	.026	.016	.016	.012	.012
		.4	.028	.028	.028	.028	.027	.027	.020	.018	.013	.013
		.8	.033	.029	.033	.029	.025	.028	.019	.020	.014	.015
g-h-2	6 bis 12	.0	.032	.034	.032	.034	.031	.032	.025	.026	.017	.017
		.4	.033	.036	.033	.036	.032	.031	.027	.026	.019	.017
		.8	.037	.035	.037	.037	.031	.031	.028	.025	.024	.025
	13 bis 20	.0	.027	.028	.027	.028	.028	.026	.016	.016	.012	.012
		.4	.029	.029	.029	.030	.027	.027	.018	.019	.013	.015
		.8	.033	.035	.033	.035	.031	.031	.025	.025	.016	.010
g-h-3	6 bis 12	.0	.032	.033	.032	.033	.032	.032	.025	.024	.017	.017
		.4	.035	.035	.035	.035	.031	.032	.026	.026	.019	.019
		.8	.037	.035	.037	.035	.033	.032	.028	.027	.028	.020
	13 bis 20	.0	.027	.026	.027	.026	.028	.027	.016	.015	.013	.013
		.4	.029	.030	.029	.030	.027	.028	.019	.018	.012	.014
		.8	.033	.031	.035	.031	.027	.028	.022	.021	.013	.012
g-h-4	6 bis 12	.0	.032	.034	.032	.034	.031	.032	.025	.026	.017	.016
		.4	.033	.037	.034	.037	.032	.031	.027	.026	.019	.019
		.8	.038	.037	.038	.038	.034	.032	.032	.026	.027	.020
	13 bis 20	.0	.028	.027	.028	.028	.029	.027	.016	.017	.012	.012
		.4	.028	.029	.028	.029	.027	.027	.018	.019	.012	.014
		.8	.031	.032	.031	.033	.027	.033	.022	.026	.020	.024
Hyper	6 bis 12	.0	.038	.037	.038	.037	.037	.037	.025	.026	.017	.019
		.4	.042	.041	.043	.041	.039	.040	.030	.032	.020	.024
		.8	.047	.044	.047	.044	.045	.043	.039	.037	.029	.034
	13 bis 20	.0	.035	.034	.035	.034	.035	.034	.019	.021	.017	.018
		.4	.038	.037	.038	.038	.037	.037	.025	.025	.021	.023
		.8	.046	.041	.046	.041	.044	.041	.038	.034	.030	.031

Anmerkung: Vorderspalte 1 gibt die Verteilung an, Vorderspalte 2 den Stichprobenumfang N , Vorderspalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer (siehe Kapitel 3.2.3) aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; bei kleineren Stichproben basiert jede Zelle auf 35 Simulationsbedingungen, bei größeren Stichproben sind es 40 Bedingungen je Zelle, siehe hierzu das Untersuchungsdesign in Kapitel 3.2.4

Tabelle F-2. Mittlere Intervalllänge

			MAX		Pratt		Blyth		MID-P		Wilson	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	6 bis 12	.0	.570	.570	.572	.572	.557	.558	.526	.526	.482	.482
		.4	.543	.556	.544	.558	.536	.547	.503	.514	.461	.471
		.8	.480	.535	.481	.536	.486	.530	.452	.496	.411	.454
	13 bis 20	.0	.428	.428	.428	.428	.422	.422	.393	.393	.375	.375
		.4	.402	.414	.402	.415	.401	.411	.369	.381	.354	.364
		.8	.336	.393	.336	.394	.346	.394	.313	.362	.300	.347
Gleich	6 bis 12	.0	.570	.570	.571	.572	.557	.557	.525	.526	.482	.482
		.4	.537	.552	.538	.553	.531	.543	.498	.510	.456	.467
		.8	.474	.522	.475	.523	.481	.519	.447	.487	.406	.444
	13 bis 20	.0	.428	.428	.428	.428	.422	.422	.393	.393	.375	.375
		.4	.396	.410	.396	.410	.396	.407	.364	.377	.349	.360
		.8	.330	.379	.331	.379	.341	.382	.308	.350	.295	.335
g-h-1	6 bis 12	.0	.570	.570	.572	.572	.557	.558	.526	.526	.482	.482
		.4	.546	.555	.547	.557	.538	.546	.505	.513	.463	.470
		.8	.488	.533	.489	.534	.493	.528	.458	.494	.418	.453
	13 bis 20	.0	.428	.427	.428	.428	.422	.422	.393	.393	.375	.375
		.4	.404	.413	.405	.414	.403	.410	.372	.380	.356	.363
		.8	.345	.392	.346	.392	.354	.392	.320	.360	.308	.345
g-h-2	6 bis 12	.0	.570	.570	.572	.572	.557	.558	.526	.526	.482	.482
		.4	.545	.540	.547	.542	.538	.534	.505	.501	.463	.459
		.8	.496	.490	.497	.491	.499	.494	.464	.461	.424	.419
	13 bis 20	.0	.428	.427	.428	.428	.422	.422	.393	.393	.375	.375
		.4	.404	.398	.405	.399	.402	.398	.371	.366	.355	.351
		.8	.354	.345	.354	.346	.361	.354	.327	.322	.315	.308
g-h-3	6 bis 12	.0	.570	.570	.571	.571	.557	.557	.525	.525	.482	.482
		.4	.544	.552	.545	.553	.537	.543	.504	.510	.461	.468
		.8	.486	.524	.487	.526	.491	.521	.457	.487	.416	.446
	13 bis 20	.0	.427	.428	.428	.428	.422	.422	.393	.393	.375	.375
		.4	.403	.410	.403	.411	.402	.408	.370	.377	.355	.360
		.8	.343	.383	.344	.383	.352	.385	.318	.352	.306	.338
g-h-4	6 bis 12	.0	.570	.570	.572	.572	.558	.558	.526	.526	.482	.482
		.4	.542	.539	.544	.541	.536	.533	.502	.500	.460	.458
		.8	.489	.487	.490	.489	.493	.492	.459	.459	.418	.417
	13 bis 20	.0	.428	.427	.429	.428	.422	.422	.393	.393	.375	.375
		.4	.401	.397	.402	.398	.400	.397	.369	.365	.353	.350
		.8	.346	.343	.346	.344	.355	.353	.321	.320	.308	.306
Hyper	6 bis 12	.0	.573	.573	.575	.575	.560	.560	.516	.516	.485	.484
		.4	.549	.554	.551	.556	.541	.545	.494	.498	.466	.470
		.8	.492	.528	.493	.529	.496	.524	.446	.473	.422	.450
	13 bis 20	.0	.429	.430	.430	.430	.423	.424	.388	.389	.376	.376
		.4	.406	.411	.407	.411	.404	.408	.366	.370	.357	.361
		.8	.347	.384	.347	.384	.355	.386	.312	.344	.309	.339

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer (siehe Kapitel 3.2.3) aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; bei kleineren Stichproben basiert jede Zelle auf 35 Simulationsbedingungen, bei größeren Stichproben sind es 40 Bedingungen je Zelle, siehe hierzu das Untersuchungsdesign in Kapitel 3.2.4

Tabelle F-3. $|CP - .95|$

			Agresti		Agresti(No)		Cliff-Wald		Cliff-Halperin		Cliff-Logit	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	6 bis 12	.0	.019	.018	.019	.018	.087	.086	.019	.018	.029	.030
		.4	.020	.019	.020	.019	.071	.082	.020	.020	.028	.029
		.8	.025	.019	.025	.019	.070	.075	.020	.020	.023	.027
	13 bis 20	.0	.013	.013	.013	.013	.034	.036	.015	.014	.019	.019
		.4	.014	.014	.014	.014	.071	.051	.016	.016	.019	.021
		.8	.024	.014	.024	.014	.055	.067	.023	.015	.025	.021
Gleich	6 bis 12	.0	.018	.018	.018	.018	.086	.087	.019	.019	.029	.029
		.4	.019	.020	.019	.020	.082	.069	.019	.020	.027	.029
		.8	.026	.026	.026	.026	.069	.064	.020	.028	.023	.031
	13 bis 20	.0	.016	.014	.016	.014	.037	.039	.016	.014	.019	.020
		.4	.015	.014	.015	.014	.080	.069	.014	.017	.019	.023
		.8	.024	.016	.024	.016	.064	.057	.023	.017	.024	.021
g-h-1	6 bis 12	.0	.019	.019	.019	.019	.087	.085	.019	.019	.028	.029
		.4	.020	.020	.020	.020	.075	.085	.020	.019	.028	.029
		.8	.024	.019	.024	.019	.070	.084	.028	.022	.029	.027
	13 bis 20	.0	.013	.013	.013	.013	.034	.037	.015	.013	.019	.018
		.4	.014	.014	.014	.014	.064	.052	.015	.015	.020	.021
		.8	.016	.017	.016	.017	.051	.077	.014	.016	.018	.022
g-h-2	6 bis 12	.0	.019	.019	.019	.019	.086	.086	.020	.019	.030	.030
		.4	.020	.018	.020	.018	.086	.075	.021	.020	.027	.028
		.8	.022	.027	.022	.027	.076	.072	.024	.025	.028	.029
	13 bis 20	.0	.013	.013	.013	.013	.034	.037	.015	.014	.019	.019
		.4	.014	.016	.014	.016	.062	.073	.015	.015	.021	.021
		.8	.018	.022	.018	.022	.068	.033	.017	.012	.020	.017
g-h-3	6 bis 12	.0	.019	.018	.019	.018	.087	.087	.019	.019	.030	.030
		.4	.020	.020	.020	.020	.070	.073	.020	.019	.028	.031
		.8	.024	.020	.024	.020	.070	.084	.027	.021	.028	.027
	13 bis 20	.0	.015	.014	.015	.014	.039	.039	.014	.013	.020	.019
		.4	.013	.015	.013	.015	.067	.059	.015	.016	.019	.021
		.8	.017	.015	.017	.015	.052	.073	.013	.015	.017	.019
g-h-4	6 bis 12	.0	.019	.018	.019	.018	.085	.086	.020	.019	.029	.029
		.4	.019	.019	.019	.019	.074	.074	.020	.018	.028	.027
		.8	.024	.026	.024	.026	.078	.069	.027	.018	.028	.024
	13 bis 20	.0	.013	.013	.013	.013	.033	.036	.015	.014	.020	.020
		.4	.013	.016	.013	.016	.070	.077	.015	.015	.019	.019
		.8	.019	.024	.019	.024	.046	.044	.018	.025	.020	.029
Hyper	6 bis 12	.0	.021	.022	.012	.013	.078	.075	.015	.018	.023	.026
		.4	.026	.028	.011	.016	.072	.066	.016	.016	.022	.024
		.8	.038	.036	.017	.032	.055	.054	.013	.019	.018	.023
	13 bis 20	.0	.021	.020	.008	.008	.042	.041	.010	.011	.015	.016
		.4	.024	.026	.009	.007	.060	.045	.012	.011	.014	.017
		.8	.038	.034	.013	.016	.049	.044	.015	.011	.016	.016

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer (siehe Kapitel 3.2.3) aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; bei kleineren Stichproben basiert jede Zelle auf 35 Simulationsbedingungen, bei größeren Stichproben sind es 40 Bedingungen je Zelle, siehe hierzu das Untersuchungsdesign in Kapitel 3.2.4

Tabelle F-4. Mittlere Intervalllänge

			Agresti		Agresti(No)		Cliff-Wald		Cliff-Halperin		Cliff-Logit	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	6 bis 12	.0	.501	.501	.501	.501	.517	.517	.500	.500	.547	.547
		.4	.484	.492	.484	.492	.466	.491	.474	.487	.525	.536
		.8	.442	.479	.442	.479	.352	.451	.411	.466	.463	.517
	13 bis 20	.0	.386	.385	.386	.385	.398	.398	.384	.384	.406	.406
		.4	.367	.376	.367	.376	.358	.378	.361	.372	.386	.396
		.8	.320	.361	.320	.361	.261	.346	.296	.353	.326	.379
Gleich	6 bis 12	.0	.501	.501	.501	.501	.516	.517	.499	.500	.547	.547
		.4	.480	.489	.480	.489	.454	.482	.468	.482	.519	.532
		.8	.438	.469	.438	.469	.340	.429	.405	.452	.458	.502
	13 bis 20	.0	.386	.385	.386	.385	.398	.398	.384	.384	.406	.406
		.4	.363	.373	.363	.373	.349	.371	.355	.368	.382	.393
		.8	.316	.350	.316	.350	.252	.325	.290	.337	.321	.363
g-h-1	6 bis 12	.0	.501	.501	.501	.501	.517	.518	.500	.500	.547	.548
		.4	.486	.492	.486	.492	.471	.488	.477	.486	.527	.535
		.8	.448	.477	.448	.477	.365	.446	.420	.465	.473	.516
	13 bis 20	.0	.386	.385	.386	.385	.398	.398	.384	.384	.406	.406
		.4	.369	.376	.369	.376	.363	.376	.363	.371	.389	.396
		.8	.327	.361	.327	.361	.274	.343	.306	.352	.337	.379
g-h-2	6 bis 12	.0	.501	.501	.501	.501	.517	.518	.500	.500	.547	.548
		.4	.485	.482	.485	.482	.470	.461	.477	.471	.527	.521
		.8	.453	.448	.453	.448	.379	.371	.428	.419	.482	.470
	13 bis 20	.0	.386	.385	.386	.385	.398	.398	.384	.384	.406	.406
		.4	.369	.365	.369	.365	.362	.353	.363	.357	.388	.383
		.8	.333	.326	.333	.326	.287	.276	.315	.303	.346	.331
g-h-3	6 bis 12	.0	.501	.500	.501	.500	.516	.516	.499	.499	.547	.547
		.4	.484	.489	.484	.489	.467	.482	.475	.483	.526	.532
		.8	.446	.472	.446	.472	.362	.430	.418	.456	.471	.508
	13 bis 20	.0	.385	.385	.385	.385	.398	.398	.384	.384	.406	.406
		.4	.368	.373	.368	.373	.360	.371	.362	.368	.387	.393
		.8	.326	.354	.326	.354	.271	.330	.304	.343	.336	.371
g-h-4	6 bis 12	.0	.501	.501	.501	.501	.517	.517	.500	.500	.547	.548
		.4	.484	.481	.484	.481	.464	.459	.474	.470	.525	.521
		.8	.448	.446	.448	.446	.367	.367	.421	.418	.474	.469
	13 bis 20	.0	.386	.385	.386	.385	.399	.398	.384	.384	.407	.406
		.4	.367	.364	.367	.364	.358	.352	.360	.356	.386	.383
		.8	.327	.325	.327	.325	.275	.272	.306	.302	.337	.332
Hyper	6 bis 12	.0	.503	.503	.476	.479	.493	.495	.474	.472	.515	.513
		.4	.488	.491	.454	.458	.438	.446	.441	.439	.483	.479
		.8	.450	.475	.405	.432	.317	.383	.365	.395	.407	.433
	13 bis 20	.0	.387	.387	.365	.367	.376	.379	.362	.362	.381	.380
		.4	.371	.374	.343	.346	.334	.340	.332	.331	.352	.348
		.8	.328	.356	.291	.319	.231	.288	.256	.288	.277	.304

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer (siehe Kapitel 3.2.3) aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; bei kleineren Stichproben basiert jede Zelle auf 35 Simulationsbedingungen, bei größeren Stichproben sind es 40 Bedingungen je Zelle, siehe hierzu das Untersuchungsdesign in Kapitel 3.2.4

Tabelle F-5. $|\overline{CP} - .95|$

			CLI-WALD-Z		CLI-WALD-T		CLI-HAL-Z		CLI-HAL-T		CLI-LOG-Z		CLI-LOG-T	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	6 bis 12	.0	.089	.095	.057	.070	.048	.061	.023	.038	.032	.047	.020	.036
		.4	.079	.087	.048	.059	.053	.057	.024	.041	.041	.045	.018	.033
		.8	.108	.086	.074	.059	.092	.064	.059	.045	.085	.054	.050	.034
	13 bis 20	.0	.041	.043	.023	.028	.013	.017	.007	.017	.005	.013	.012	.017
		.4	.032	.037	.014	.021	.013	.016	.006	.016	.007	.013	.010	.017
		.8	.035	.033	.019	.020	.025	.017	.012	.017	.022	.014	.009	.016
Gleich	6 bis 12	.0	.083	.088	.049	.060	.040	.050	.017	.032	.024	.039	.018	.031
		.4	.078	.079	.043	.051	.049	.050	.023	.032	.039	.041	.018	.031
		.8	.112	.081	.077	.052	.097	.060	.065	.041	.090	.052	.059	.039
	13 bis 20	.0	.037	.040	.020	.024	.010	.011	.007	.011	.003	.009	.014	.015
		.4	.028	.033	.011	.017	.010	.010	.007	.013	.005	.010	.012	.016
		.8	.032	.031	.016	.016	.025	.014	.013	.015	.022	.013	.011	.016
g-h-1	6 bis 12	.0	.098	.099	.066	.072	.060	.065	.033	.047	.042	.050	.026	.038
		.4	.087	.088	.057	.061	.062	.062	.033	.045	.050	.050	.023	.036
		.8	.114	.090	.080	.063	.097	.069	.061	.047	.088	.057	.050	.035
	13 bis 20	.0	.044	.044	.028	.029	.018	.019	.011	.018	.010	.014	.012	.018
		.4	.036	.037	.019	.022	.017	.017	.009	.017	.011	.013	.009	.018
		.8	.039	.032	.024	.022	.028	.019	.015	.018	.024	.015	.011	.017
g-h-2	6 bis 12	.0	.094	.079	.064	.050	.059	.038	.029	.021	.043	.025	.029	.021
		.4	.088	.071	.058	.040	.066	.040	.038	.024	.054	.032	.029	.020
		.8	.113	.084	.080	.053	.101	.060	.060	.043	.093	.054	.055	.037
	13 bis 20	.0	.043	.039	.026	.022	.016	.010	.010	.008	.007	.005	.013	.016
		.4	.035	.028	.018	.012	.016	.008	.011	.008	.009	.005	.012	.014
		.8	.036	.029	.021	.015	.025	.017	.012	.013	.021	.015	.009	.013
g-h-3	6 bis 12	.0	.097	.101	.067	.073	.060	.064	.029	.039	.042	.049	.027	.037
		.4	.090	.086	.059	.058	.064	.058	.034	.040	.052	.045	.024	.032
		.8	.117	.088	.083	.060	.101	.066	.058	.036	.093	.055	.053	.031
	13 bis 20	.0	.044	.044	.028	.029	.018	.017	.012	.015	.010	.012	.013	.017
		.4	.035	.037	.018	.021	.017	.016	.009	.015	.011	.011	.009	.016
		.8	.039	.032	.025	.020	.028	.018	.015	.015	.024	.013	.011	.014
g-h-4	6 bis 12	.0	.086	.077	.058	.047	.054	.035	.026	.021	.038	.023	.028	.022
		.4	.083	.070	.053	.039	.062	.040	.036	.025	.051	.032	.028	.021
		.8	.113	.087	.080	.057	.101	.067	.072	.049	.096	.062	.062	.049
	13 bis 20	.0	.040	.037	.023	.020	.012	.008	.008	.009	.004	.004	.013	.017
		.4	.032	.027	.015	.010	.012	.007	.009	.010	.006	.005	.011	.015
		.8	.036	.028	.019	.015	.026	.018	.013	.015	.022	.017	.010	.015
Hyper	6 bis 12	.0	.093	.082	.060	.054	.056	.048	.031	.031	.037	.039	.026	.029
		.4	.083	.073	.050	.045	.058	.044	.032	.029	.045	.039	.022	.028
		.8	.107	.076	.075	.050	.092	.056	.060	.040	.083	.051	.049	.033
	13 bis 20	.0	.043	.035	.026	.020	.017	.012	.011	.016	.009	.013	.013	.020
		.4	.034	.030	.018	.015	.017	.012	.009	.014	.011	.011	.010	.017
		.8	.038	.030	.023	.015	.029	.016	.016	.013	.025	.015	.013	.013

Anmerkung: Vorderspalte 1 gibt die Verteilung an, Vorderspalte 2 den Stichprobenumfang N , Vorderspalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer (siehe Kapitel 3.2.3) aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; bei kleineren Stichproben basiert jede Zelle auf 35 Simulationsbedingungen, bei größeren Stichproben sind es 40 Bedingungen je Zelle, siehe hierzu das Untersuchungsdesign in Kapitel 3.2.4

Tabelle F-6. Mittlere Intervalllänge

			CLI-WALD-Z		CLI-WALD-T		CLI-HAL-Z		CLI-HAL-T		CLI-LOG-Z		CLI-LOG-T	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	6 bis 12	.0	.429	.461	.494	.528	.405	.433	.465	.495	.432	.466	.507	.544
		.4	.359	.425	.418	.491	.346	.405	.402	.465	.366	.433	.433	.508
		.8	.261	.383	.306	.445	.260	.370	.306	.428	.272	.393	.326	.464
	13 bis 20	.0	.325	.353	.351	.381	.311	.338	.334	.362	.321	.351	.347	.380
		.4	.267	.323	.289	.348	.260	.311	.280	.335	.266	.323	.288	.349
		.8	.192	.290	.207	.313	.190	.282	.205	.304	.193	.291	.210	.315
Gleich	6 bis 12	.0	.426	.463	.490	.530	.401	.434	.461	.496	.428	.467	.502	.545
		.4	.350	.430	.408	.496	.337	.407	.392	.468	.355	.435	.420	.510
		.8	.242	.388	.284	.451	.241	.373	.284	.431	.251	.396	.302	.467
	13 bis 20	.0	.321	.354	.347	.382	.306	.337	.329	.362	.316	.351	.341	.379
		.4	.258	.326	.280	.352	.251	.312	.270	.336	.257	.323	.278	.349
		.8	.174	.293	.188	.317	.172	.283	.186	.305	.175	.292	.189	.316
g-h-1	6 bis 12	.0	.434	.458	.499	.525	.410	.431	.470	.493	.439	.464	.514	.541
		.4	.359	.413	.418	.478	.348	.395	.404	.455	.369	.422	.437	.495
		.8	.268	.367	.314	.427	.269	.358	.316	.415	.284	.380	.339	.449
	13 bis 20	.0	.331	.351	.357	.378	.317	.336	.341	.361	.329	.351	.355	.379
		.4	.269	.313	.291	.338	.263	.304	.284	.327	.270	.315	.293	.341
		.8	.200	.278	.217	.300	.199	.273	.216	.294	.204	.282	.221	.305
g-h-2	6 bis 12	.0	.438	.445	.503	.511	.414	.417	.475	.478	.444	.446	.520	.522
		.4	.366	.389	.426	.452	.355	.369	.412	.427	.377	.391	.446	.461
		.8	.275	.329	.322	.385	.276	.317	.324	.370	.292	.333	.348	.395
	13 bis 20	.0	.334	.336	.360	.363	.320	.319	.344	.343	.333	.330	.359	.357
		.4	.276	.288	.298	.312	.270	.277	.291	.298	.278	.284	.301	.307
		.8	.207	.241	.223	.261	.206	.234	.223	.253	.212	.239	.230	.259
g-h-3	6 bis 12	.0	.435	.453	.500	.520	.411	.427	.471	.488	.440	.459	.515	.536
		.4	.355	.403	.413	.467	.345	.386	.401	.446	.366	.412	.433	.484
		.8	.266	.356	.311	.415	.267	.348	.313	.404	.281	.368	.336	.436
	13 bis 20	.0	.331	.348	.357	.375	.317	.333	.341	.358	.329	.347	.356	.374
		.4	.267	.305	.288	.329	.261	.296	.281	.318	.268	.306	.290	.331
		.8	.198	.269	.214	.291	.198	.264	.214	.285	.202	.272	.219	.294
g-h-4	6 bis 12	.0	.438	.445	.503	.511	.414	.416	.475	.477	.443	.445	.519	.521
		.4	.369	.393	.429	.456	.357	.371	.414	.429	.379	.393	.448	.463
		.8	.274	.327	.321	.383	.275	.314	.323	.365	.290	.330	.347	.390
	13 bis 20	.0	.333	.335	.359	.362	.319	.318	.343	.342	.331	.329	.358	.355
		.4	.278	.291	.300	.315	.271	.279	.292	.300	.279	.286	.302	.309
		.8	.206	.238	.222	.258	.205	.231	.221	.249	.210	.235	.228	.254
Hyper	6 bis 12	.0	.428	.465	.494	.532	.404	.435	.464	.496	.431	.467	.506	.545
		.4	.356	.437	.415	.504	.345	.411	.401	.471	.364	.438	.431	.514
		.8	.257	.404	.301	.468	.258	.383	.303	.441	.270	.407	.323	.478
	13 bis 20	.0	.325	.356	.351	.384	.312	.337	.335	.361	.323	.350	.349	.378
		.4	.267	.331	.289	.358	.261	.314	.281	.338	.271	.325	.293	.351
		.8	.192	.304	.208	.329	.192	.291	.207	.312	.195	.299	.212	.323

Anmerkung: Vorpalte 1 gibt die Verteilung an, Vorpalte 2 den Stichprobenumfang N , Vorpalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer (siehe Kapitel 3.2.3) aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; bei kleineren Stichproben basiert jede Zelle auf 35 Simulationsbedingungen, bei größeren Stichproben sind es 40 Bedingungen je Zelle, siehe hierzu das Untersuchungsdesign in Kapitel 3.2.4

Tabelle F-7. $|\overline{CP} - .95|$

			MRT-WALD		RT-WALD		MRT-HAL		RT-HAL		MRT-LOG		RT-LOG	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	6 bis 12	.0	.016	.015	.026	.015	.028	.024	.037	.035	.033	.031	.041	.040
		.4	.009	.022	.022	.008	.015	.006	.030	.023	.019	.010	.033	.029
		.8	.028	.094	.027	.049	.024	.053	.024	.018	.022	.044	.023	.016
	13 bis 20	.0	.016	.013	.022	.009	.024	.014	.030	.023	.028	.020	.034	.028
		.4	.010	.029	.019	.014	.015	.010	.024	.008	.018	.007	.026	.011
		.8	.023	.109	.023	.079	.017	.079	.019	.050	.015	.073	.018	.044
Gleich	6 bis 12	.0	.017	.019	.028	.015	.027	.024	.036	.035	.032	.031	.040	.041
		.4	.011	.030	.025	.011	.016	.007	.030	.021	.020	.010	.032	.028
		.8	.016	.097	.020	.054	.015	.058	.023	.023	.014	.049	.023	.020
	13 bis 20	.0	.019	.018	.025	.013	.024	.014	.031	.022	.028	.019	.034	.027
		.4	.014	.034	.022	.021	.017	.017	.026	.011	.020	.014	.028	.012
		.8	.008	.112	.014	.083	.007	.086	.015	.058	.007	.080	.015	.052
g-h-1	6 bis 12	.0	.014	.013	.024	.015	.027	.024	.037	.036	.032	.031	.041	.040
		.4	.006	.023	.017	.008	.010	.006	.027	.022	.014	.008	.030	.028
		.8	.052	.103	.043	.057	.038	.060	.030	.021	.035	.051	.028	.019
	13 bis 20	.0	.011	.010	.018	.007	.022	.015	.029	.023	.026	.020	.033	.028
		.4	.003	.031	.010	.016	.009	.011	.019	.006	.012	.006	.022	.009
		.8	.054	.119	.047	.087	.037	.087	.031	.056	.034	.080	.029	.049
g-h-2	6 bis 12	.0	.012	.019	.023	.022	.025	.027	.036	.037	.031	.033	.040	.041
		.4	.006	.020	.013	.015	.010	.010	.026	.025	.014	.014	.030	.029
		.8	.055	.061	.045	.027	.034	.043	.027	.020	.031	.037	.026	.019
	13 bis 20	.0	.010	.020	.017	.018	.020	.020	.028	.028	.025	.025	.031	.032
		.4	.006	.026	.007	.018	.007	.017	.018	.016	.010	.017	.021	.018
		.8	.054	.070	.047	.048	.036	.059	.030	.036	.033	.055	.028	.033
g-h-3	6 bis 12	.0	.013	.013	.023	.016	.027	.025	.037	.036	.033	.031	.041	.041
		.4	.006	.021	.016	.007	.009	.006	.026	.022	.012	.008	.029	.027
		.8	.059	.100	.049	.054	.041	.060	.032	.023	.038	.053	.030	.021
	13 bis 20	.0	.011	.012	.018	.009	.022	.016	.029	.025	.027	.021	.033	.029
		.4	.004	.028	.009	.013	.007	.010	.017	.006	.010	.006	.020	.009
		.8	.059	.116	.051	.085	.040	.087	.036	.057	.038	.081	.033	.050
g-h-4	6 bis 12	.0	.012	.022	.023	.023	.025	.026	.035	.036	.031	.032	.040	.041
		.4	.006	.022	.016	.018	.012	.013	.027	.027	.016	.017	.031	.031
		.8	.044	.048	.037	.018	.027	.031	.026	.016	.025	.025	.026	.016
	13 bis 20	.0	.011	.023	.018	.021	.021	.019	.028	.027	.025	.024	.031	.031
		.4	.005	.028	.010	.023	.009	.021	.019	.019	.013	.020	.022	.021
		.8	.043	.052	.037	.032	.029	.045	.024	.025	.026	.041	.022	.023
Hyper	6 bis 12	.0	.012	.027	.023	.017	.027	.021	.036	.033	.032	.028	.040	.039
		.4	.004	.043	.019	.020	.012	.013	.028	.020	.016	.012	.031	.026
		.8	.054	.136	.051	.083	.034	.084	.032	.035	.032	.073	.031	.028
	13 bis 20	.0	.011	.025	.018	.019	.022	.019	.029	.021	.026	.020	.033	.024
		.4	.004	.047	.012	.032	.010	.031	.020	.021	.013	.026	.023	.019
		.8	.051	.143	.047	.112	.034	.117	.032	.085	.031	.110	.029	.077

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer (siehe Kapitel 3.2.3) aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; bei kleineren Stichproben basiert jede Zelle auf 35 Simulationsbedingungen, bei größeren Stichproben sind es 40 Bedingungen je Zelle, siehe hierzu das Untersuchungsdesign in Kapitel 3.2.4

Tabelle F-8. Mittlere Intervalllänge

			MRT-WALD		RT-WALD		MRT-HAL		RT-HAL		MRT-LOG		RT-LOG	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	6 bis 12	.0	.530	.526	.577	.576	.536	.542	.576	.583	.606	.612	.653	.661
		.4	.439	.446	.488	.500	.446	.462	.486	.507	.494	.517	.541	.569
		.8	.296	.326	.334	.384	.306	.343	.338	.392	.331	.380	.367	.434
	13 bis 20	.0	.374	.371	.393	.392	.370	.372	.387	.390	.395	.400	.416	.422
		.4	.301	.308	.319	.328	.298	.309	.314	.327	.312	.327	.330	.347
		.8	.192	.219	.204	.239	.192	.221	.203	.239	.196	.229	.208	.248
Gleich	6 bis 12	.0	.531	.526	.578	.575	.534	.540	.573	.581	.603	.611	.649	.659
		.4	.436	.443	.483	.496	.440	.458	.479	.501	.487	.511	.532	.562
		.8	.288	.325	.326	.379	.295	.338	.326	.383	.317	.371	.351	.421
	13 bis 20	.0	.374	.372	.394	.392	.369	.370	.387	.389	.394	.397	.414	.419
		.4	.298	.307	.316	.327	.294	.306	.310	.324	.307	.322	.325	.342
		.8	.186	.218	.197	.236	.185	.218	.196	.234	.188	.224	.200	.242
g-h-1	6 bis 12	.0	.529	.526	.577	.576	.539	.542	.579	.583	.608	.612	.656	.661
		.4	.426	.433	.477	.487	.437	.450	.479	.496	.485	.503	.534	.556
		.8	.281	.307	.322	.365	.295	.326	.330	.375	.322	.361	.360	.416
	13 bis 20	.0	.373	.371	.393	.392	.371	.373	.389	.391	.398	.402	.419	.423
		.4	.291	.298	.309	.318	.290	.299	.307	.318	.305	.317	.323	.337
		.8	.183	.206	.196	.225	.184	.208	.196	.226	.189	.216	.202	.235
g-h-2	6 bis 12	.0	.529	.529	.577	.577	.540	.535	.579	.575	.609	.605	.656	.653
		.4	.428	.433	.479	.483	.441	.439	.483	.480	.490	.486	.539	.534
		.8	.287	.304	.332	.349	.303	.312	.341	.348	.332	.336	.373	.376
	13 bis 20	.0	.373	.373	.393	.393	.371	.369	.389	.387	.399	.394	.419	.415
		.4	.293	.297	.311	.316	.293	.293	.310	.310	.309	.306	.327	.324
		.8	.188	.200	.202	.214	.190	.198	.203	.211	.196	.202	.210	.216
g-h-3	6 bis 12	.0	.529	.526	.577	.575	.539	.541	.579	.582	.609	.612	.656	.661
		.4	.419	.426	.470	.480	.431	.442	.473	.487	.478	.493	.527	.545
		.8	.274	.299	.315	.354	.288	.315	.324	.363	.315	.348	.354	.399
	13 bis 20	.0	.373	.372	.393	.392	.371	.372	.389	.390	.398	.400	.419	.422
		.4	.286	.292	.304	.312	.285	.293	.302	.311	.300	.309	.318	.329
		.8	.178	.199	.191	.217	.179	.200	.191	.217	.184	.207	.197	.224
g-h-4	6 bis 12	.0	.529	.529	.577	.577	.538	.535	.578	.574	.607	.604	.655	.651
		.4	.436	.443	.486	.492	.447	.447	.487	.487	.496	.495	.543	.541
		.8	.297	.316	.339	.359	.310	.321	.346	.356	.338	.345	.377	.384
	13 bis 20	.0	.373	.374	.393	.393	.371	.369	.389	.387	.397	.394	.418	.414
		.4	.299	.305	.317	.323	.298	.300	.315	.316	.313	.313	.331	.331
		.8	.194	.207	.207	.221	.195	.205	.207	.218	.200	.209	.213	.222
Hyper	6 bis 12	.0	.521	.512	.569	.562	.527	.524	.568	.567	.595	.592	.643	.643
		.4	.425	.429	.476	.482	.433	.441	.474	.485	.480	.490	.528	.542
		.8	.273	.301	.312	.356	.281	.309	.315	.356	.306	.337	.342	.389
	13 bis 20	.0	.368	.364	.388	.384	.364	.360	.382	.379	.389	.385	.410	.406
		.4	.293	.299	.311	.319	.292	.296	.309	.314	.306	.310	.324	.330
		.8	.180	.206	.192	.224	.179	.204	.191	.221	.184	.208	.196	.226

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer (siehe Kapitel 3.2.3) aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; bei kleineren Stichproben basiert jede Zelle auf 35 Simulationsbedingungen, bei größeren Stichproben sind es 40 Bedingungen je Zelle, siehe hierzu das Untersuchungsdesign in Kapitel 3.2.4

Tabelle F-9. $|\overline{CP} - .95|$

			S-CLI-Z-1		S-CLI-T-1		S-CLI-Z-2		S-CLI-T-2		S-CLI-Z-3		S-CLI-T-3	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	6 bis 12	.0	.024	.021	.011	.015	.014	.009	.018	.020	.016	.012	.015	.018
		.4	.040	.037	.011	.006	.031	.023	.019	.013	.032	.028	.013	.009
		.8	.069	.066	.028	.018	.068	.055	.040	.024	.067	.058	.028	.018
	13 bis 20	.0	.007	.009	.009	.009	.003	.004	.013	.013	.004	.006	.012	.011
		.4	.012	.018	.007	.003	.007	.011	.011	.006	.009	.014	.008	.004
		.8	.024	.032	.011	.012	.015	.024	.010	.007	.019	.028	.010	.009
Gleich	6 bis 12	.0	.023	.027	.011	.010	.014	.013	.017	.016	.016	.018	.014	.014
		.4	.040	.050	.011	.010	.035	.035	.018	.012	.035	.040	.013	.009
		.8	.077	.085	.037	.040	.081	.070	.046	.043	.079	.076	.034	.038
	13 bis 20	.0	.005	.012	.011	.006	.003	.007	.015	.009	.003	.009	.013	.007
		.4	.011	.024	.006	.006	.006	.018	.012	.003	.008	.021	.009	.004
		.8	.024	.045	.014	.025	.018	.036	.013	.019	.021	.040	.013	.021
g-h-1	6 bis 12	.0	.022	.018	.013	.018	.013	.009	.019	.022	.014	.010	.017	.021
		.4	.035	.029	.013	.010	.029	.022	.021	.016	.029	.024	.016	.013
		.8	.062	.057	.020	.013	.062	.049	.035	.022	.060	.050	.022	.015
	13 bis 20	.0	.007	.007	.010	.011	.003	.003	.014	.014	.005	.005	.012	.012
		.4	.012	.014	.007	.005	.006	.008	.011	.009	.008	.011	.009	.007
		.8	.024	.028	.009	.007	.014	.021	.010	.003	.019	.024	.009	.005
g-h-2	6 bis 12	.0	.017	.018	.018	.016	.009	.013	.022	.020	.010	.011	.021	.019
		.4	.035	.036	.010	.015	.029	.034	.017	.019	.030	.033	.012	.015
		.8	.065	.069	.015	.044	.065	.076	.031	.050	.063	.075	.015	.042
	13 bis 20	.0	.006	.006	.012	.013	.004	.006	.015	.016	.005	.005	.013	.014
		.4	.013	.016	.007	.010	.007	.014	.011	.011	.009	.015	.009	.010
		.8	.023	.034	.006	.027	.015	.030	.005	.024	.019	.032	.005	.025
g-h-3	6 bis 12	.0	.021	.018	.014	.017	.013	.009	.020	.022	.014	.011	.017	.020
		.4	.035	.032	.014	.008	.030	.022	.022	.016	.030	.024	.016	.012
		.8	.064	.059	.019	.014	.063	.051	.034	.025	.061	.052	.021	.016
	13 bis 20	.0	.007	.007	.010	.011	.003	.003	.014	.014	.005	.004	.012	.013
		.4	.011	.014	.007	.005	.006	.008	.012	.009	.008	.011	.010	.007
		.8	.024	.026	.008	.007	.015	.019	.009	.006	.019	.023	.008	.006
g-h-4	6 bis 12	.0	.019	.019	.016	.014	.010	.014	.019	.018	.012	.013	.019	.017
		.4	.038	.037	.006	.016	.031	.034	.015	.019	.033	.034	.009	.015
		.8	.071	.072	.022	.047	.073	.080	.040	.053	.072	.080	.023	.045
	13 bis 20	.0	.007	.006	.011	.012	.005	.006	.015	.016	.006	.006	.013	.014
		.4	.014	.014	.006	.011	.009	.013	.010	.012	.011	.013	.008	.011
		.8	.029	.033	.013	.028	.020	.031	.010	.025	.024	.032	.011	.026
Hyper	6 bis 12	.0	.136	.186	.066	.129	.025	.029	.036	.039	.011	.021	.017	.013
		.4	.185	.236	.103	.178	.022	.019	.029	.030	.025	.052	.013	.018
		.8	.245	.309	.170	.259	.023	.026	.022	.025	.050	.107	.021	.061
	13 bis 20	.0	.156	.178	.107	.151	.033	.037	.039	.041	.003	.009	.014	.010
		.4	.234	.207	.182	.182	.033	.030	.038	.036	.006	.031	.010	.015
		.8	.223	.231	.200	.212	.034	.014	.038	.017	.017	.064	.004	.046

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer (siehe Kapitel 3.2.3) aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; bei kleineren Stichproben basiert jede Zelle auf 35 Simulationsbedingungen, bei größeren Stichproben sind es 40 Bedingungen je Zelle, siehe hierzu das Untersuchungsdesign in Kapitel 3.2.4

Tabelle F-10. Mittlere Intervalllänge

			S-CLI-Z-1		S-CLI-T-1		S-CLI-Z-2		S-CLI-T-2		S-CLI-Z-3		S-CLI-T-3	
			1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1	1:1	3:1
Normal	6 bis 12	.0	.399	.423	.462	.488	.426	.450	.486	.514	.413	.438	.474	.501
		.4	.333	.383	.394	.448	.355	.407	.410	.467	.345	.396	.402	.457
		.8	.244	.338	.296	.399	.257	.355	.299	.407	.249	.347	.299	.403
	13 bis 20	.0	.309	.333	.333	.358	.317	.340	.341	.365	.313	.337	.337	.362
		.4	.255	.303	.276	.326	.263	.310	.283	.333	.259	.307	.279	.329
		.8	.183	.270	.197	.290	.190	.277	.204	.297	.186	.273	.201	.294
Gleich	6 bis 12	.0	.395	.423	.458	.488	.422	.451	.482	.513	.410	.438	.471	.501
		.4	.324	.384	.384	.447	.345	.407	.399	.466	.335	.396	.392	.456
		.8	.226	.340	.276	.399	.237	.355	.277	.405	.230	.347	.277	.402
	13 bis 20	.0	.306	.334	.330	.358	.314	.341	.337	.366	.310	.337	.333	.362
		.4	.248	.305	.267	.328	.255	.312	.275	.335	.251	.309	.271	.332
		.8	.166	.273	.180	.294	.173	.280	.187	.300	.170	.276	.184	.297
g-h-1	6 bis 12	.0	.403	.422	.468	.488	.430	.449	.492	.513	.418	.437	.480	.500
		.4	.334	.374	.395	.439	.355	.397	.411	.457	.345	.386	.404	.448
		.8	.252	.327	.307	.389	.264	.343	.308	.395	.257	.335	.309	.392
	13 bis 20	.0	.315	.331	.338	.356	.322	.339	.346	.364	.318	.335	.342	.360
		.4	.257	.294	.277	.317	.264	.302	.284	.324	.260	.298	.281	.321
		.8	.190	.259	.205	.279	.197	.266	.212	.286	.194	.263	.209	.283
g-h-2	6 bis 12	.0	.405	.419	.471	.486	.432	.444	.494	.510	.419	.433	.482	.498
		.4	.338	.369	.400	.437	.358	.387	.414	.449	.348	.378	.407	.443
		.8	.257	.313	.313	.380	.268	.318	.311	.371	.261	.314	.313	.376
	13 bis 20	.0	.316	.324	.340	.349	.324	.331	.348	.356	.320	.328	.344	.353
		.4	.262	.281	.282	.304	.269	.288	.290	.311	.266	.284	.286	.307
		.8	.196	.235	.212	.256	.203	.242	.219	.261	.200	.238	.215	.258
g-h-3	6 bis 12	.0	.404	.420	.468	.487	.431	.447	.493	.511	.418	.435	.481	.499
		.4	.330	.369	.392	.435	.351	.392	.406	.452	.341	.381	.400	.444
		.8	.250	.322	.305	.386	.261	.337	.304	.388	.255	.330	.306	.387
	13 bis 20	.0	.315	.330	.339	.355	.322	.337	.346	.362	.318	.334	.342	.358
		.4	.254	.289	.274	.311	.262	.296	.282	.319	.258	.293	.278	.315
		.8	.188	.253	.203	.273	.195	.260	.210	.280	.192	.257	.207	.276
g-h-4	6 bis 12	.0	.404	.420	.469	.487	.430	.445	.492	.511	.418	.433	.481	.499
		.4	.339	.373	.401	.442	.360	.390	.415	.453	.350	.382	.408	.447
		.8	.255	.316	.310	.383	.265	.315	.308	.369	.259	.312	.310	.376
	13 bis 20	.0	.315	.324	.339	.349	.323	.331	.347	.357	.319	.328	.343	.353
		.4	.263	.284	.284	.308	.271	.291	.291	.314	.267	.287	.288	.311
		.8	.196	.234	.211	.255	.203	.240	.218	.259	.199	.237	.215	.257
Hyper	6 bis 12	.0	.313	.328	.375	.387	.521	.548	.578	.607	.416	.435	.477	.496
		.4	.248	.287	.306	.341	.452	.498	.506	.553	.349	.389	.406	.446
		.8	.173	.243	.215	.285	.359	.435	.400	.478	.260	.332	.305	.379
	13 bis 20	.0	.223	.246	.249	.268	.412	.431	.436	.455	.316	.335	.340	.359
		.4	.161	.218	.184	.237	.350	.397	.373	.419	.263	.302	.283	.324
		.8	.118	.192	.126	.207	.297	.359	.308	.377	.193	.266	.208	.284

Anmerkung: Vorspalte 1 gibt die Verteilung an, Vorspalte 2 den Stichprobenumfang N , Vorspalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer (siehe Kapitel 3.2.3) aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; bei kleineren Stichproben basiert jede Zelle auf 35 Simulationsbedingungen, bei größeren Stichproben sind es 40 Bedingungen je Zelle, siehe hierzu das Untersuchungsdesign in Kapitel 3.2.4

Tabelle F-11. $|\overline{CP} - .95|$

			S-MRT-1		S-MRT-2		S-MRT-3	
			1:1	3:1	1:1	3:1	1:1	3:1
Normal	6 bis 12	.0	.022	.015	.029	.023	.026	.019
		.4	.012	.015	.021	.010	.017	.012
		.8	.024	.095	.006	.060	.011	.076
	13 bis 20	.0	.021	.011	.024	.013	.023	.012
		.4	.013	.021	.018	.017	.015	.019
		.8	.020	.107	.010	.093	.014	.100
Gleich	6 bis 12	.0	.022	.014	.029	.022	.026	.018
		.4	.013	.026	.022	.018	.017	.021
		.8	.015	.109	.011	.073	.007	.088
	13 bis 20	.0	.023	.014	.025	.014	.024	.014
		.4	.015	.030	.020	.025	.018	.027
		.8	.005	.117	.004	.103	.002	.109
g-h-1	6 bis 12	.0	.021	.015	.028	.024	.025	.020
		.4	.007	.015	.017	.009	.012	.009
		.8	.042	.100	.013	.062	.024	.078
	13 bis 20	.0	.019	.011	.022	.014	.021	.012
		.4	.007	.020	.011	.014	.009	.017
		.8	.045	.117	.033	.101	.038	.109
g-h-2	6 bis 12	.0	.020	.019	.027	.026	.024	.023
		.4	.006	.021	.017	.019	.011	.019
		.8	.043	.070	.016	.048	.027	.056
	13 bis 20	.0	.018	.018	.021	.020	.019	.019
		.4	.005	.025	.009	.023	.006	.023
		.8	.047	.072	.038	.059	.042	.065
g-h-3	6 bis 12	.0	.021	.016	.028	.025	.025	.021
		.4	.006	.014	.017	.009	.011	.009
		.8	.047	.100	.016	.062	.029	.077
	13 bis 20	.0	.019	.012	.022	.015	.021	.014
		.4	.005	.018	.010	.014	.007	.016
		.8	.050	.113	.037	.097	.043	.105
g-h-4	6 bis 12	.0	.019	.019	.027	.026	.024	.023
		.4	.007	.024	.018	.022	.013	.023
		.8	.034	.057	.013	.044	.020	.050
	13 bis 20	.0	.018	.020	.021	.021	.020	.020
		.4	.007	.028	.011	.027	.009	.027
		.8	.038	.056	.029	.051	.033	.053
Hyper	6 bis 12	.0	.042	.099	.041	.038	.026	.023
		.4	.092	.182	.037	.028	.015	.038
		.8	.247	.357	.028	.020	.021	.133
	13 bis 20	.0	.082	.147	.041	.037	.022	.023
		.4	.178	.224	.038	.030	.011	.044
		.8	.280	.332	.035	.028	.036	.144

Anmerkung: Vorderspalte 1 gibt die Verteilung an, Vorderspalte 2 den Stichprobenumfang N , Vorderspalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer (siehe Kapitel 3.2.3) aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; bei kleineren Stichproben basiert jede Zelle auf 35 Simulationsbedingungen, bei größeren Stichproben sind es 40 Bedingungen je Zelle, siehe hierzu das Untersuchungsdesign in Kapitel 3.2.4

Tabelle F-12. Mittlere Intervalllänge

			S-MRT-1		S-MRT-2		S-MRT-3	
			1:1	3:1	1:1	3:1	1:1	3:1
Normal	6 bis 12	.0	.509	.507	.536	.534	.523	.521
		.4	.420	.426	.448	.454	.434	.440
		.8	.279	.305	.310	.335	.295	.320
	13 bis 20	.0	.362	.361	.369	.368	.366	.364
		.4	.292	.299	.299	.306	.296	.302
		.8	.186	.212	.194	.219	.190	.215
Gleich	6 bis 12	.0	.510	.508	.537	.535	.523	.521
		.4	.416	.423	.445	.452	.430	.437
		.8	.272	.303	.303	.333	.287	.318
	13 bis 20	.0	.363	.361	.370	.368	.366	.365
		.4	.289	.298	.297	.305	.293	.302
		.8	.180	.211	.187	.218	.184	.214
g-h-1	6 bis 12	.0	.508	.507	.535	.534	.522	.521
		.4	.407	.413	.436	.442	.422	.428
		.8	.265	.287	.296	.318	.281	.303
	13 bis 20	.0	.362	.360	.369	.368	.365	.364
		.4	.282	.289	.290	.296	.286	.293
		.8	.177	.199	.185	.206	.181	.202
g-h-2	6 bis 12	.0	.508	.509	.535	.536	.522	.522
		.4	.409	.415	.437	.444	.423	.430
		.8	.272	.288	.302	.318	.287	.303
	13 bis 20	.0	.361	.362	.369	.369	.365	.365
		.4	.284	.289	.292	.296	.288	.293
		.8	.182	.194	.190	.202	.186	.198
g-h-3	6 bis 12	.0	.509	.507	.535	.534	.522	.520
		.4	.401	.407	.430	.436	.415	.422
		.8	.259	.280	.289	.311	.274	.296
	13 bis 20	.0	.361	.361	.369	.368	.365	.364
		.4	.278	.284	.285	.291	.281	.288
		.8	.172	.193	.180	.200	.176	.196
g-h-4	6 bis 12	.0	.508	.509	.535	.536	.522	.522
		.4	.417	.425	.445	.453	.431	.439
		.8	.281	.300	.312	.330	.297	.315
	13 bis 20	.0	.362	.362	.369	.369	.365	.365
		.4	.290	.296	.298	.303	.294	.300
		.8	.188	.201	.196	.209	.192	.205
Hyper	6 bis 12	.0	.422	.409	.605	.607	.515	.508
		.4	.326	.322	.519	.525	.423	.422
		.8	.178	.196	.380	.393	.278	.286
	13 bis 20	.0	.274	.270	.452	.446	.361	.358
		.4	.193	.207	.377	.381	.289	.293
		.8	.104	.134	.284	.292	.181	.199

Anmerkung: Vorpalte 1 gibt die Verteilung an, Vorpalte 2 den Stichprobenumfang N , Vorpalte 3 die Korrelation ρ ; in der Kopfzeile sind die untersuchten Schätzer (siehe Kapitel 3.2.3) aufgeführt, Zeile 2 bezeichnet das Verhältnis $\sigma_1:\sigma_2$; bei kleineren Stichproben basiert jede Zelle auf 35 Simulationsbedingungen, bei größeren Stichproben sind es 40 Bedingungen je Zelle, siehe hierzu das Untersuchungsdesign in Kapitel 3.2.4