

Vergleich der Edgington-Methode mit den Standard-Verfahren für Fixed-Effect- Metaanalysen im Fall von 2 Studien

A decorative horizontal bar spanning the width of the page, composed of 18 small squares in various shades of blue and grey. The word 'ARBEITSPAPIER' is centered in white text on a dark blue rectangular background that covers the middle portion of this bar.

ARBEITSPAPIER

Projekt: GA25-06

Version: 1.0

Stand: 17.07.2026

IQWiG-Berichte – Nr. 2289

DOI: 10.60584/GA25-06

Impressum

Herausgeber

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Thema

Vergleich der Edgington-Methode mit den Standard-Verfahren für Fixed-Effect-Metaanalysen im Fall von 2 Studien

Auftraggeber

Bearbeitung im Rahmen des Generalauftrags

Interne Projektnummer

GA25-06

DOI-URL

<https://doi.org/10.60584/GA25-06>

Anschrift des Herausgebers

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
Siegburger Str. 237
50679 Köln

Tel.: +49 221 35685-0
Fax: +49 221 35685-1
E-Mail: info@iqwig.de
Internet: www.iqwig.de

ISSN: 1864-2500

Zitiervorschlag

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Vergleich der Edgington-Methode mit den Standard-Verfahren für Fixed-Effect-Metaanalysen im Fall von 2 Studien; Arbeitspapier [online]. 2026 [Zugriff: TT.MM.JJJJ]. URL: <https://doi.org/10.60584/GA25-06>.

Schlagwörter

Modelle – statistische, Metaanalyse, Nutzenbewertung, Heterogenität, wenige Studien

Keywords

Models – Statistical, Meta-Analysis as Topic, Benefit Assessment, Heterogeneity, Sparse Data

Dieses Arbeitspapier wurde unter Beteiligung externer Sachverständiger erstellt. Externe Sachverständige, die wissenschaftliche Forschungsaufträge für das Institut bearbeiten, haben gemäß § 139b Abs. 3 Satz 2 Sozialgesetzbuch – Fünftes Buch – Gesetzliche Krankenversicherung „alle Beziehungen zu Interessenverbänden, Auftragsinstituten, insbesondere der pharmazeutischen Industrie und der Medizinprodukteindustrie, einschließlich Art und Höhe von Zuwendungen“ offenzulegen. Das Institut hat von allen Sachverständigen ein ausgefülltes „Formblatt zur Offenlegung von Beziehungen“ erhalten. Die Angaben wurden durch das speziell für die Beurteilung der Interessenkonflikte eingerichtete Gremium des Instituts bewertet. Die Selbstangaben der externen Sachverständigen und der externen Revierinnen und Reviewer zur Offenlegung von Beziehungen sind in Anhang A dargestellt. Es wurden keine Interessenkonflikte festgestellt, die die fachliche Unabhängigkeit im Hinblick auf eine Bearbeitung des vorliegenden Auftrags gefährden.

Für die Inhalte des Berichts ist allein das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) verantwortlich.

Externe Sachverständige

- Leonhard Held (Institut für Epidemiologie, Biostatistik und Prävention (EBPI) der Universität Zürich, Zürich)
- Samuel Pawel (Institut für Epidemiologie, Biostatistik und Prävention (EBPI) der Universität Zürich, Zürich)

Das IQWiG dankt den externen Beteiligten für ihre Mitarbeit am Projekt.

Mitarbeiterinnen und Mitarbeiter des IQWiG

- Leonie Schürmeyer
- Ralf Bender
- Guido Skipka
- Sibylle Sturtz
- Frank Weber

Inhaltsverzeichnis

	Seite
Tabellenverzeichnis	v
Abbildungsverzeichnis	vi
Abkürzungsverzeichnis.....	vii
Kurzfassung	viii
1 Hintergrund.....	1
2 Fragestellung.....	3
3 Projektverlauf.....	4
4 Methoden	5
4.1 Edgington-Methode.....	6
4.2 Gewichtete Versionen der Edgington-Methode.....	7
4.3 Software-Nutzung	8
5 Ergebnisse	9
5.1 Charakterisierung der IQWiG-Metaanalysen-Datenbank	9
5.2 Vergleich der Konfidenzintervalle	10
5.3 Vergleich der statistischen Signifikanz.....	11
6 Diskussion	15
7 Fazit	17
8 Literatur	18
Anhang A Offenlegung von Beziehungen der externen Sachverständigen	21

Tabellenverzeichnis

	Seite
Tabelle 1: Anzahl Metaanalysen in der IQWiG-Datenbank (getrennt nach Effektmaß) mit prozentualen Anteilen der statistisch signifikanten FE-Metaanalysen und Q-Tests (jeweils zum Niveau $\alpha = 0,05$).....	9
Tabelle 2: Charakteristika der 95 %-Konfidenzintervalle (95 %-KIs) der Metaanalysen aus der IQWiG-Metaanalysen-Datenbank, getrennt nach verwendeter Methode	10

Abbildungsverzeichnis

	Seite
Abbildung 1: Beispiel für einen Drapery Plot (oben) mit dem zugehörigen Forest Plot (unten) bei einer Metaanalyse mit 2 Studien.....	7
Abbildung 2: Prozentuale Anteile der Übereinstimmung bzw. Nicht-Übereinstimmung bezüglich der Signifikanzaussage zwischen der kombinierten FE-Methodik IV + MH einerseits und allen betrachteten Versionen der Edgington-Methode sowie der RE-Methodik andererseits	12
Abbildung 3: Beispiele aus der IQWiG-Metaanalysen-Datenbank, bei denen mindestens 1 Version der Edgington-Methode zu einer anderen Signifikanzaussage als die FE-Methodik IV + MH führt.....	13

Abkürzungsverzeichnis

Abkürzung	Bedeutung
DL	DerSimonian-Laird
FE	Fixed Effect
G-BA	Gemeinsamer Bundesausschuss
HK	Hartung-Knapp
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
IV	Inverse Varianz
KI	Konfidenzintervall
MH	Mantel-Haenszel
RCT	Randomized controlled Trial (randomisierte kontrollierte Studie)
RE	Random Effects
SE	Standard Error (Standardfehler)

Kurzfassung

Im Rahmen des Generalauftrags wurde das Thema Vergleich der Edgington-Methode mit den Standard-Verfahren für Fixed-Effect-Metaanalysen im Fall von 2 Studien bearbeitet.

Fragestellung

Das übergeordnete Ziel des Projekts ist die Beantwortung der Frage, ob die bisherigen Verfahren für Metaanalysen mit 2 Studien sinnvoll durch andere Methoden ersetzt werden können. In diesem Kooperationsprojekt mit dem Departement Biostatistik am Institut für Epidemiologie, Biostatistik und Prävention der Universität Zürich soll die Edgington-Methode mit den bisher eingesetzten Standard-Verfahren für FE-Metaanalysen (Inverse-Varianz [IV] und Mantel-Haenszel [MH]) empirisch mithilfe von Metaanalysen aus einer Datenbank des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) verglichen werden, um zu klären, ob sich die Anwendung der Edgington-Methode für Metaanalysen mit 2 Studien bereits aus empirischer Sicht als nicht sinnvoll erweist (z. B. aufgrund eines systematisch zu breiten KIs für den Gesamteffekt).

Methoden

Basierend auf den studienspezifischen Daten aus der zuvor aufbereiteten IQWiG-Metaanalysen-Datenbank wurde ein empirischer Vergleich der Verfahren durchgeführt. Neben der Edgington-Methode (in ihrer ungewichteten Version sowie in 2 gewichteten Versionen) und den bisher am IQWiG standardmäßig eingesetzten Fixed-Effect(FE)-Verfahren wurden zur besseren Einordnung der Ergebnisse auch übliche Random-Effects(RE)-Verfahren betrachtet.

Ergebnisse

Insgesamt wurden 1226 Metaanalysen mit 2 Studien in die Analyse eingeschlossen. Die Edgington-Methode (sowohl in der ungewichteten Version als auch in den 2 gewichteten Versionen) zeigte dabei gute Eigenschaften in Hinblick auf die Breite des zugehörigen Konfidenzintervalls (KIs) sowie in Hinblick auf die Übereinstimmung (mit den bisher am IQWiG genutzten FE-Verfahren) in der Signifikanzaussage des entsprechenden statistischen Tests. Insbesondere sind die Edgington-KIs in allen untersuchten Anwendungen informativ und in der deutlichen Mehrheit der Fälle schmaler als mindestens ein studienspezifisches KI. Anhand von Beispielen aus der IQWiG-Datenbank wurde demonstriert, wie das KI der Edgington-Methode auf verschiedene Datensituationen reagieren kann. So kann das KI der ungewichteten Edgington-Methode bei hoher Diskrepanz der studienspezifischen Punkt- und /oder Variabilitätsschätzungen sehr breit ausfallen. Um die studienspezifischen Variabilitätsschätzungen weiter zu berücksichtigen, lassen sich die gewichteten Versionen der

Edgington-Methode verwenden. In den vorgestellten Beispielen war zu erkennen, dass die Gewichtung anhand der inversen Varianzen die metaanalytische Punktschätzung stark in Richtung der Punktschätzung der präziseren Studie verschiebt, während die Gewichtung anhand der inversen Standardfehler einen Mittelweg zwischen der ungewichteten Edgington-Methode und der Edgington-Methode mit Gewichtung anhand der inversen Varianzen darstellt.

Fazit

Basierend auf der vorliegenden empirischen Untersuchung der Edgington-Methode für Metaanalysen bei 2 Studien anhand der Datenbank vergangener IQWiG-Metaanalysen könnte die Edgington-Methode eine sinnvolle (möglicherweise sogar besser geeignete) Alternative zu den aktuell am IQWiG im Fall von 2 Studien standardmäßig genutzten FE-Verfahren sein.

Um schlussendlich entscheiden zu können, ob für IQWiG-Metaanalysen mit 2 Studien zukünftig standardmäßig die Edgington-Methode zum Einsatz kommen sollte, bedarf es einer weiteren Untersuchung in Form einer Simulationsstudie. Die Ergebnisse der und Schlussfolgerungen aus der Simulationsstudie werden in einem separaten, noch zu erstellenden Arbeitspapier dargestellt.

1 Hintergrund

Um die Evidenz aus mehreren Studien zusammenfassen zu können, werden Metaanalysen durchgeführt. Anhand dieser soll der zu untersuchende Behandlungseffekt präziser geschätzt werden als auf Basis der jeweils einzelnen Studien. Im Allgemeinen werden 2 grundlegende Arten von Metaanalysen unterschieden: Metaanalysen auf Basis eines festen (gemeinsamen) Effekts (Fixed-Effect[FE]- oder Common-Effect-Metaanalysen) sowie Metaanalysen auf Basis zufälliger Effekte (Random-Effects[RE]-Metaanalysen). Bei FE-Metaanalysen wird angenommen, dass alle Studien einen gemeinsamen wahren Effekt schätzen und sich studienspezifische Unterschiede lediglich durch die Stichprobenvariabilität innerhalb der Studien ergeben. Bei RE-Metaanalysen wird hingegen angenommen, dass sich der wahre Effekt je nach Studie unterscheiden kann und dieser einer übergeordneten Verteilung entstammt, wobei die zugehörige Varianz dieser Verteilung den Grad der Heterogenität zwischen den Studien widerspiegelt [1-3].

Auch bei Metaanalysen von nur 2 Studien wäre es prinzipiell erstrebenswert, die Heterogenität zwischen den Studien in die metaanalytische Zusammenfassung einzubeziehen, allerdings birgt dies große Herausforderungen. Im Falle von nur 2 Studien kann die Varianz zwischen den Studien nicht zuverlässig geschätzt werden, was ein wichtiger Grund ist, weshalb Metaanalysen mit zufälligen Effekten in diesem Fall instabile bzw. in der Praxis untaugliche Ergebnisse liefern. Die resultierenden Konfidenzintervalle (KIs) können übermäßig breit ausfallen und Schätzungen für die Heterogenität sind nicht verlässlich [4-6].

Die aktuelle Methode zur Durchführung von Metaanalysen im IQWiG [7] sieht deshalb bei 2 Studien vor, eine FE-Metaanalyse durchzuführen. Bei 3 bis 4 Studien soll eine bayessche RE-Metaanalyse durchgeführt werden, welche insbesondere auf Vorwissen zum Grad der Heterogenität basiert (das Vorwissen über den wahren mittleren Behandlungseffekt wird als „nichtinformativ“ angenommen – widerspiegelt durch eine vollkommen flache, degenerierte Verteilung). Ab 5 Studien wird die Knapp-Hartung-Methode (auch als Hartung-Knapp[HK]-Methode bezeichnet) zur Durchführung einer Metaanalyse mit zufälligen Effekten verwendet [8,9]. Die Verwendung einer FE-Metaanalyse für 2 Studien stellt hierbei eine rein pragmatische Lösung dar, die nicht zufriedenstellend ist, aber gewählt wurde, da bisher kein Verfahren für eine RE-Metaanalyse gefunden wurde, das regelhaft für alle Effektmaße in dieser Situation gut funktioniert.

Neue Untersuchungen [10,11] deuten auf gute Eigenschaften der sogenannten Edgington-Methode hin, ein spezielles Verfahren für die Berechnung von p-Wert-Funktionen (auch Konfidenzkurven [12] oder Konfidenzverteilungen [13] genannt), das auch für Metaanalysen nutzbar ist. Bisherige Simulationen im Fall von mehr als 2 Studien [10] zeigen, dass die Edgington-Methode in heterogenen Situationen bessere Eigenschaften hat als die Standard-Verfahren für FE-Metaanalysen. Die KIs der Edgington-Methode sind nicht notwendigerweise

symmetrisch und dadurch kann dieses Verfahren besser auf Diskrepanz der studienspezifischen Effektschätzungen reagieren als die üblichen Standard-Verfahren für FE-Metaanalysen.

2 Fragestellung

Das übergeordnete Ziel des Projekts ist die Beantwortung der Frage, ob die bisherigen Verfahren für Metaanalysen mit 2 Studien sinnvoll durch andere Methoden ersetzt werden können. In diesem Kooperationsprojekt mit dem Departement Biostatistik am Institut für Epidemiologie, Biostatistik und Prävention der Universität Zürich soll die Edgington-Methode mit den bisher eingesetzten Standard-Verfahren für FE-Metaanalysen (Inverse-Varianz [IV] und Mantel-Haenszel [MH]) empirisch mithilfe von Metaanalysen aus einer Datenbank des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) verglichen werden, um zu klären, ob sich die Anwendung der Edgington-Methode für Metaanalysen mit 2 Studien bereits aus empirischer Sicht als nicht sinnvoll erweist (z. B. aufgrund eines systematisch zu breiten KIs für den Gesamteffekt).

3 Projektverlauf

Im Rahmen des Generalauftrags wurde der empirische Vergleich der Edgington-Methode mit den Standard-Verfahren für FE-Metaanalysen im Fall von 2 Studien bearbeitet. Bearbeitungsbeginn war der 15.11.2025.

Auf Basis einer internen Projektskizze wurde ein Arbeitspapier erstellt. Dieser Bericht wurde an den G-BA übermittelt und 4 Wochen später auf der Website des IQWiG veröffentlicht.

Zusätzlich zum Arbeitspapier ist auch eine Publikation in einer Fachzeitschrift geplant, in der weitere technische Details enthalten sind.

4 Methoden

Zunächst wurde eine Aufbereitung der IQWiG-Metaanalysen-Datenbank durchgeführt. Anschließend wurden vorhandene R-Programme der Züricher Kooperationspartner auf die Daten aus der IQWiG-Metaanalysen-Datenbank angewendet, um den empirischen Vergleich der Verfahren durchzuführen.

Für diesen empirischen Vergleich (bei Metaanalysen mit 2 Studien) wurden Standard-Verfahren aus den Bereichen der FE- und RE-Metaanalysen sowie die Edgington-Methode betrachtet, letztere in der klassischen (ungewichteten) und in 2 gewichteten Versionen.

Im Folgenden wird die im Arbeitspapier genutzte Notation und die zentral behandelte Edgington-Methode sowie 2 Erweiterungen dieser beschrieben. Hierzu wird angenommen, dass eine Metaanalyse für die Zusammenfassung der Evidenz aus 2 Studien durchgeführt wird. Dabei bezeichnet $\hat{\theta}_i$ den Schätzer des Effekts θ_i sowie σ_i den zugehörigen Standardfehler für die Studie $i \in \{1; 2\}$. In der Praxis wird σ_i durch eine entsprechende Schätzung ersetzt, es wird also (wie in Metaanalysen üblich) angenommen, dass σ_i perfekt geschätzt wird. Für die Nullhypothese $H_{0i}: \theta_i = \mu$ gegenüber der Alternativhypothese $H_{1i}: \theta_i > \mu$ bei einem gegebenen $\mu \in (-\infty; \infty)$ ergibt sich unter der Annahme einer approximativen Normalverteilung des Schätzers für den zugehörigen einseitigen p-Wert die Formel

$$p_i(\mu) = 1 - \Phi \left\{ \frac{(\hat{\theta}_i - \mu)}{\sigma_i} \right\},$$

wobei $\Phi(\cdot)$ der Verteilungsfunktion der Standardnormalverteilung entspricht. Wird $p_i(\mu)$ als Funktion von μ betrachtet, so ergibt sich eine sogenannte p-Wert-Funktion, auch bekannt als Konfidenzkurve [12] oder Konfidenzverteilung [13].

Es gibt verschiedene Methoden, mit denen p-Wert-Funktionen aus verschiedenen Studien zu einer metaanalytischen p-Wert-Funktion zusammengefasst werden können, aus der wiederum metaanalytische Punkt- und Intervallschätzungen für den Gesamteffekt berechnet werden können [13]. Bei vielen dieser Methoden hängt das Ergebnis von der Orientierung der Alternativhypothese ab, d. h. die Ergebnisse für die Alternativhypothesen $H_{1i}: \theta_i > \mu$ und $\tilde{H}_{1i}: \theta_i < \mu$ sind im Allgemeinen unterschiedlich. Im Unterschied dazu ist die Edgington-Methode invariant gegenüber der Orientierung der Alternativhypothese [10]. Basierend auf dieser Eigenschaft, weiteren günstigen theoretischen Eigenschaften [11] und guten Ergebnissen in einer bereits durchgeführten Simulationsstudie für Metaanalysen mit mindestens 3 Studien [10] wurde die Edgington-Methode für den Vergleich von metaanalytischen Verfahren bei 2 Studien für diesen Generalauftrag ausgewählt.

4.1 Edgington-Methode

Die klassische (ungewichtete) Edgington-Methode basiert im Falle von 2 Studien auf der ungewichteten Summe der studienspezifischen p-Wert-Funktionen: $s(\mu) = p_1(\mu) + p_2(\mu)$. Allgemein gilt, dass ein p-Wert unter der Nullhypothese gleichverteilt auf dem Intervall $[0; 1]$ ist. Wird angenommen, dass alle studienspezifischen Nullhypothesen $H_{0i}: \theta_i = \mu$ wahr sind, so folgt $s(\mu)$ einer Irwin-Hall-Verteilung mit dem Parameter $n = 2$ [14,15], welche einer Dreiecksverteilung entspricht. An der Annahme, dass alle studienspezifischen Nullhypothesen wahr sind, lässt sich erkennen, dass die Edgington-Methode die FE-Annahme zugrunde legt, auch wenn in Kapitel 5 zu sehen sein wird, dass die Edgington-Methode dennoch auf Diskrepanzen in den Studienergebnissen reagieren kann. Sofern nicht anders beschrieben, sind daher mit „FE-Verfahren“ (oder „FE-Methoden“ etc.) im Folgenden nur solche FE-Verfahren gemeint, die bereits etabliert sind und nicht der Edgington-Methode entsprechen.

Basierend auf der Verteilungsfunktion der oben genannten Irwin-Hall-Verteilung kann eine metaanalytische Zusammenfassung der p-Werte anhand der Edgington-Methode mit der kombinierten p-Wert-Funktion

$$p_E(\mu) = \begin{cases} \frac{\{s(\mu)\}^2}{2}, & \text{wenn } 0 \leq s(\mu) \leq 1 \\ 1 - \frac{\{2 - s(\mu)\}^2}{2}, & \text{wenn } 1 < s(\mu) \leq 2 \end{cases}$$

durchgeführt werden. Anhand der kombinierten p-Wert-Funktion $p_E(\mu)$ kann eine metaanalytische Punktschätzung für den Gesamteffekt durch Auflösen der Gleichung $p_E(\mu) = \frac{1}{2}$ nach μ , sowie ein $(1 - \alpha) \times 100$ %-KI für den Gesamteffekt durch das Lösen der Gleichungen $p_E(\mu) = \frac{\alpha}{2}$ und $p_E(\mu) = 1 - \frac{\alpha}{2}$ berechnet werden. Im Fall von 2 Studien ergibt sich für die

Punktschätzung die geschlossene Form $\hat{\mu} = \frac{\frac{\hat{\theta}_1 + \hat{\theta}_2}{\frac{1}{\sigma_1} + \frac{1}{\sigma_2}}}{\frac{1}{\sigma_1} + \frac{1}{\sigma_2}}$, wenn die oben genannte Annahme einer approximativen Normalverteilung der studienspezifischen Punktschätzer gestellt wird [11]. Diese geschlossene Form ist somit ein gewichtetes Mittel der studienspezifischen Effektschätzungen, deren Gewichte den inversen studienspezifischen Standardfehlern entsprechen. Selbst im Fall von 2 Studien lassen sich die KI-Grenzen jedoch im Allgemeinen nicht in geschlossener Form darstellen, sodass hierfür numerische Verfahren nötig sind.

Für die Veranschaulichung solcher p-Wert-Funktionen (sowohl studienspezifisch als auch metaanalytisch) ist es sinnvoll, einen sogenannten Drapery Plot zu betrachten [16]. Hierfür ist es hilfreich, die kombinierte p-Wert-Funktion $p_E(\mu)$ in eine 2-seitige Version anhand von $\tilde{p}_E(\mu) = 2 \min\{p_E(\mu), 1 - p_E(\mu)\}$ umzuwandeln [10,17]. Das Maximum dieser Funktion entspricht dann genau der Punktschätzung und die Grenzen des $(1 - \alpha) \times 100$ %-KIs können anhand der Schnittstellen mit dem y-Wert α abgelesen werden. In Abbildung 1 ist ein Beispiel

für einen Drapery Plot dargestellt. Die kombinierte p-Wert-Funktion der Edgington-Methode (blau) erreicht ihr Maximum (und damit die Punktschätzung) bei $\log \widehat{OR} = -2,83$, was in diesem Fall identisch zur Punktschätzung der MH-Methode (schwarz; dies ist eine FE-Methode) ist. Ebenso lässt sich anhand der zusätzlich dargestellten KIs im Forest Plot erkennen, dass ihre Grenzen anhand der Schnittpunkte der p-Wert-Funktionen mit der horizontalen Linie bei 0,05 abgelesen werden können.

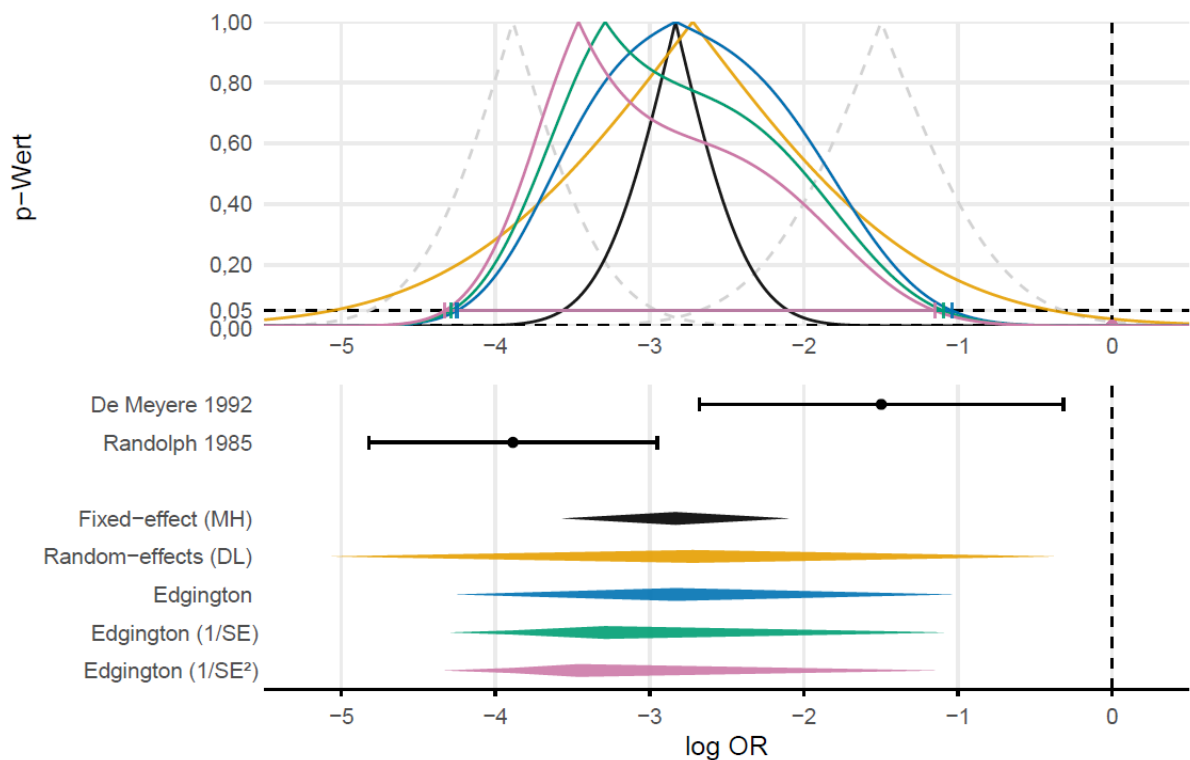


Abbildung 1: Beispiel für einen Drapery Plot (oben) mit dem zugehörigen Forest Plot (unten) bei einer Metaanalyse mit 2 Studien [18]. Gestrichelt: studienspezifische p-Wert-Funktionen. Neben der klassischen (ungewichteten) Edgington-Methode sind 2 gewichtete Erweiterungen dieser dargestellt (Gewichtung mittels inversen Standardfehlern [1/SE] und inversen Varianzen [1/SE²]), ebenso wie die MH-Methode (eine FE-Methode, vgl. [7]) und die DerSimonian-Laird(DL)-Methode (eine RE-Methode, vgl. [7]).

In Abbildung 1 sind zusätzlich zur bereits vorgestellten ungewichteten Edgington-Methode 2 gewichtete Erweiterungen dieser dargestellt, welche im Folgenden näher erläutert werden.

4.2 Gewichtete Versionen der Edgington-Methode

In der bereits vorgestellten Version der Edgington-Methode wird die gleiche Gewichtung beider Studien vorausgesetzt. Da Studien sich oft in ihrer Präzision unterscheiden, liegt es nahe, den Studien unterschiedliche Gewichte bei der Bildung der Summe der

studienspezifischen p-Wert-Funktionen zuzuweisen. Daher werden im Folgenden 2 gewichtete Erweiterungen der Edgington-Methode vorgestellt.

Im Allgemeinen lässt sich die gewichtete Summe der studienspezifischen p-Wert-Funktionen als $s_w(\mu) = w_1 p_1(\mu) + w_2 p_2(\mu)$ beschreiben, wobei w_1 und w_2 studienspezifische Gewichte bezeichnen. Ohne Beschränkung der Allgemeinheit sei $0 < w_1 \leq w_2$. Dann ist die kombinierte p-Wert-Funktion basierend auf $s_w(\mu)$ gegeben durch

$$p_w(\mu) = \begin{cases} \frac{s_w(\mu)^2}{2w_1w_2}, & \text{wenn } 0 < s_w(\mu) \leq w_1, \\ \left\{s_w(\mu) - \frac{w_1}{2}\right\}/w_1, & \text{wenn } w_1 < s_w(\mu) \leq w_2, \\ 1 - \frac{\{(w_1 + w_2) - s_w(\mu)\}^2}{2w_1w_2}, & \text{wenn } w_2 < s_w(\mu) \leq w_1 + w_2. \end{cases}$$

In diesem Arbeitspapier werden 2 verschiedene Gewichtungen genauer betrachtet: Gewichte basierend auf dem inversen Standardfehler ($w_i = 1/\sigma_i$) sowie Gewichte basierend auf der inversen Varianz ($w_i = 1/\sigma_i^2$). Das Vorgehen zur Bestimmung einer Punkt- und Intervallschätzung für den Gesamteffekt basierend auf $p_w(\mu)$ ist analog zum oben beschriebenen Vorgehen basierend auf $p_E(\mu)$, allerdings existiert bei Verwendung von $p_w(\mu)$ mit $w_1 \neq 1$ und / oder $w_2 \neq 1$ im Allgemeinen keine geschlossene Form mehr für die Punktschätzung, d. h. auch hierfür ist dann ein numerisches Verfahren nötig.

In Abbildung 1 sind auch die kombinierten p-Wert-Funktionen sowie die zugehörigen 95 %-KIs der oben genannten gewichteten Versionen der Edgington-Methode dargestellt. Es ist zu erkennen, dass die Gewichtung zu einer Verschiebung der Punktschätzung in Richtung der präziseren Studie führt, während die Grenzen des 95 %-KIs in diesem Fall nahe an denen der ungewichteten Edgington-Methode liegen.

4.3 Software-Nutzung

Für die Durchführung der Analysen wurden die statistische Programmiersprache R [19] (Version 4.6.0) sowie die Pakete confMeta [20], metafor [21], ggplot2 [22], dplyr [23], tidyr [24], ggbeeswarm [25], ggpubr [26], kableExtra [27] und knitr [28] verwendet.

5 Ergebnisse

5.1 Charakterisierung der IQWiG-Metaanalysen-Datenbank

Die Metaanalysen-Datenbank des IQWiG enthält die studienspezifischen Daten der bis einschließlich 2025 in Berichten veröffentlichten Metaanalysen seit Gründung des Instituts (für eine nähere Beschreibung des Vorgehens zur Erstellung der Metaanalysen-Datenbank, siehe [29]). Insgesamt liegen in der Datenbank studienspezifische Angaben zu 2290 Metaanalysen vor, von denen 1226 und somit über die Hälfte lediglich auf 2 Studien basieren. Alle nachfolgenden Analysen des vorliegenden Arbeitspapiers beschränken sich auf den Teil der Datenbank, der nur Metaanalysen von 2 Studien umfasst [30]. Dabei ist zu beachten, dass im Falle von zu großer Heterogenität regelhaft keine gepoolten Effektschätzungen in IQWiG-Berichten dargestellt werden. Nichtsdestotrotz werden die studienspezifischen Angaben auch bei großer Heterogenität in einem Forest Plot (ohne gepoolte Effektschätzung) dargestellt und damit in der Metaanalysen-Datenbank aufgeführt. Die in diesem Arbeitspapier durchgeführten Metaanalysen gehen deshalb über die in IQWiG-Berichten dargestellten Metaanalysen hinaus, da sie gepoolte Effektschätzungen enthalten können, die in der Praxis nicht in Berichten dargestellt würden.

Bei den betrachteten Metaanalysen wurden unterschiedliche Effektmaße zugrunde gelegt; Tabelle 1 zeigt eine Zusammenfassung der Metaanalysen getrennt nach Effektmaß.

Tabelle 1: Anzahl Metaanalysen in der IQWiG-Datenbank (getrennt nach Effektmaß) mit prozentualen Anteilen der statistisch signifikanten FE-Metaanalysen und Q-Tests (jeweils zum Niveau $\alpha = 0,05$)

Effektmaß	Metaanalysen, n (%)	FE signifikant (%)	Q-Test signifikant (%)
Logarithmiertes relatives Risiko	316 (25,8)	32,3	4,4
Logarithmiertes Odds Ratio	304 (24,8)	29,3	5,3
Mittelwertdifferenz	280 (22,8)	39,3	13,2
Standardisierte Mittelwertdifferenz	187 (15,3)	39,6	15,0
Logarithmiertes Hazard Ratio	124 (10,1)	34,7	4,8
Andere	15 (1,2)	26,7	20,0
FE: Fixed-Effect; n: Anzahl			

Am häufigsten werden logarithmierte relative Risiken und logarithmierte Odds Ratios in Metaanalysen verwendet, gefolgt von Mittelwertdifferenzen, standardisierten Mittelwertdifferenzen und logarithmierten Hazard Ratios. Andere Effektmaße kommen nur selten vor.

Mit Ausnahme der Kategorie „Andere“ zeigt sich für alle genannten Effektmaße bei Durchführung einer FE-Metaanalyse in ca. 30 % bis 40 % der Metaanalysen ein statistisch signifikanter Gesamteffekt zum Niveau $\alpha = 0,05$ sowie in höchstens 15 % ein signifikanter

Q-Test, der auf Heterogenität der Studienergebnisse hindeutet, sodass die gepoolten Effektschätzungen nicht in IQWiG-Berichten dargestellt würden.

5.2 Vergleich der Konfidenzintervalle

Um die Eigenschaften der Edgington-Methode im Vergleich zu den standardmäßig verwendeten Verfahren im Falle von Metaanalysen mit 2 Studien praxisnah zu untersuchen, wurden die Verfahren jeweils auf alle in der Datenbank vorliegenden potenziellen Metaanalysen angewandt. In Tabelle 2 sind die Eigenschaften der resultierenden 95 %-KIs aller betrachteten Methoden abgebildet. Für eine metaanalytische Zusammenfassung ist es erstrebenswert, dass das zugehörige KI nicht über die „Vereinigung“ der studienspezifischen KIs hinaus geht [6]. Wenn dies der Fall ist, wird eine Metaanalyse in diesem Sinne als informativ bezeichnet. Die erste Spalte von Tabelle 2 gibt diesen Anteil an informativen Metaanalysen für jede der betrachteten Methoden an. Der Anteil der informativen Metaanalysen aller betrachteten Edgington- und FE-Methoden ist jeweils 100 %. Dies ist für die RE-Methoden nicht der Fall: Die DerSimonian-Laird(DL)-Methode liefert in ca. 91 % der Metaanalysen ein informatives KI, während dies bei der Hartung-Knapp(HK)-Methode lediglich in unter 10 % der Metaanalysen der Fall ist. Letzteres zeigt die Unzulänglichkeit dieser Methode im Falle von nur 2 Studien besonders auf.

Tabelle 2: Charakteristika der 95 %-Konfidenzintervalle (95 %-KIs) der Metaanalysen aus der IQWiG-Metaanalysen-Datenbank, getrennt nach verwendeter Methode

Methode	Informatives KI (%)	KI schmäler als beide studienspezifischen KIs (%)	KI schmäler als ein studienspezifisches KI (%)
Edgington	100,0	62,6	29,9
Edgington (1/SE)	100,0	69,7	23,4
Edgington (1/SE ²)	100,0	76,3	17,1
Fixed-Effect (IV)	100,0	100,0	0,0
Fixed-Effect (IV + MH)	100,0	99,8	0,2
Random-Effects (DL)	90,6	72,1	12,4
Random-Effects (HK)	6,8	2,1	6,4

DL: DerSimonian-Laird; HK: Hartung-Knapp; IV: Inverse Varianz; KI: Konfidenzintervall; MH: Mantel-Haenszel; SE: Standardfehler

Die beiden letzten Spalten von Tabelle 2 geben den Anteil an Metaanalysen an, bei denen das metaanalytische KI schmäler ist als jedes einzelne studienspezifische KI bzw. schmäler ist als genau eines der beiden studienspezifischen KIs. Wünschenswert ist primär ein metaanalytisches KI, das schmäler ist als beide studienspezifischen KIs. Ist dies nicht der Fall, so ist sekundär von Interesse, ob das metaanalytische KI zumindest schmäler ist als (genau) eines der beiden studienspezifischen KIs. Die IV-Methode führt hier in allen Fällen zu einem KI, das schmäler ist als beide studienspezifischen KIs. Dies ist jedoch aufgrund der

theoretischen Eigenschaften der IV-Methode auch zu erwarten. Wird die MH-Methode bei Vorliegen einer Vierfeldertafel angewendet und ansonsten die IV-Methode (hier zusammen als IV + MH bezeichnet), so gilt diese Eigenschaft nicht mehr in allen, aber in nahezu allen untersuchten Fällen. Für die Edgington-Methode liegt der entsprechende Anteil im Bereich von 60 % bis 80 %, je nach Version der Edgington-Methode. In diesen Bereich fällt auch der Anteil bei Verwendung der DL-Methode. Im Fall der HK-Methode ist dieser Anteil äußerst klein (ca. 2 %), was die ungünstigen Eigenschaften der HK-Methode im Fall von 2 Studien unterstreicht. In Bezug auf die letzte Spalte (metaanalytisches KI schmalere als genau ein studienspezifisches KI) lässt sich erkennen, dass der Anteil bei den Edgington-Varianten gegenläufig zum Anteil aus der vorletzten Spalte (metaanalytisches KI schmalere als beide studienspezifischen KIs) steigt bzw. fällt, sodass die Summe aus diesen beiden Anteilen für die verschiedenen Versionen der Edgington-Methode stets bei etwa 93 % liegt. Im Vergleich dazu fällt der Anteil aus der letzten Spalte bei der DL-Methode geringer aus, in der Summe über die beiden Spalten ergeben sich hier nur ca. 85 %. Bei der HK-Methode fällt der Anteil der letzten Spalte sehr gering aus und ergibt auch in der Summe mit der vorletzten Spalte nur einen Wert von ca. 9 %.

5.3 Vergleich der statistischen Signifikanz

Ein weiterer Aspekt bei der Einschätzung der betrachteten Methoden ist die statistische Signifikanz. In Abbildung 2 ist die Übereinstimmung der Signifikanz aller betrachteten Methoden (auf der x-Achse) mit der aktuell am IQWiG bei 2 Studien regelhaft verwendeten FE-Methodik (IV + MH) zum Niveau $\alpha = 0,05$ dargestellt. Über alle betrachteten Methoden hinweg ist der Anteil an Metaanalysen, bei denen sowohl die jeweils betrachtete Methode auf der x-Achse als auch die kombinierte FE-Methodik IV + MH zu einem nicht statistisch signifikanten Ergebnis führen, mit etwa 65 % ähnlich hoch. Allerdings unterscheiden sich die übrigen Anteile zwischen den verschiedenen Methoden auf der x-Achse. Für den Großteil aller Metaanalysen führen IV + MH und alle Edgington-Versionen zu der gleichen Signifikanzaussage. Eine von IV + MH abweichende Signifikanzaussage liegt in ca. 9 % der Fälle für die ungewichtete Edgington-Methode vor. Für die gewichteten Edgington-Versionen fällt dieser Anteil mit ca. 8 % (Gewichtung mittels inverser Standardfehler) bzw. ca. 7 % (Gewichtung mittels inverser Varianzen) etwas geringer aus. Unterschiede in der Signifikanzaussage kommen in beiderlei Hinsicht vor: In einigen Fällen führt IV + MH zu einem statistisch signifikanten Ergebnis, während die jeweilige Edgington-Version zu einem nicht statistisch signifikanten Ergebnis führt, in anderen (weniger) Fällen verhält es sich umgekehrt. Im Gegensatz dazu führen die beiden untersuchten RE-Methoden (DL und HK) in keiner Metaanalyse, bei der die kombinierte FE-Methodik (IV + MH) zu einem nicht statistisch signifikanten Ergebnis kommt, zu statistischer Signifikanz. Des Weiteren wird bei der Verwendung der HK-Methode lediglich in ca. 3 % aller Metaanalysen ein statistisch signifikantes Ergebnis erreicht.

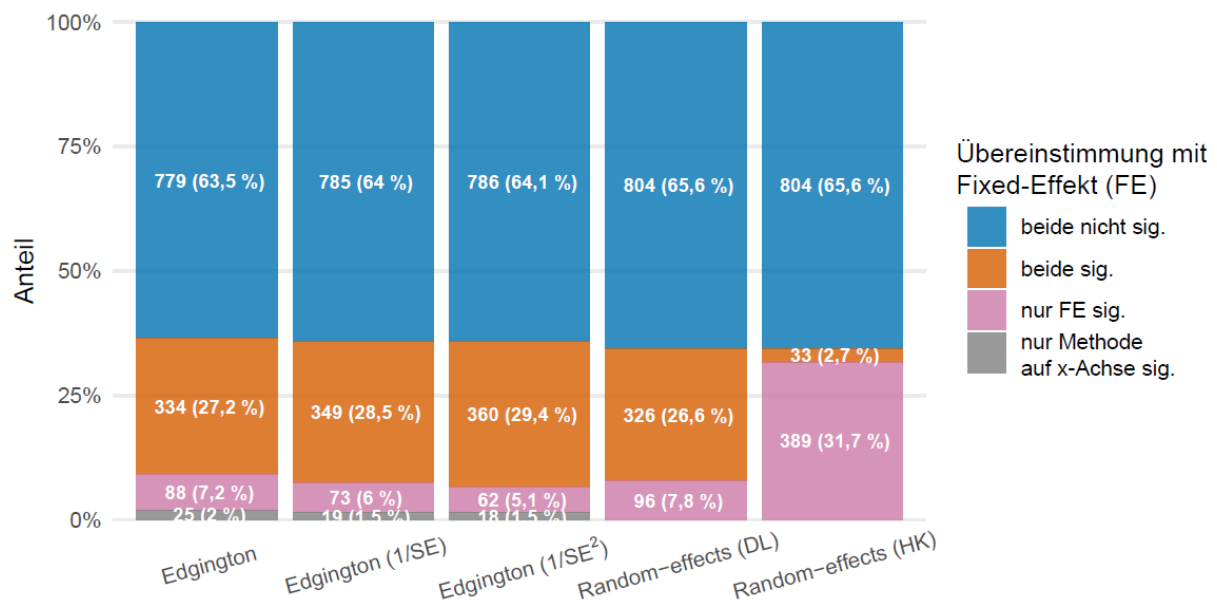


Abbildung 2: Prozentuale Anteile der Übereinstimmung bzw. Nicht-Übereinstimmung bezüglich der Signifikanzaussage zwischen der kombinierten FE-Methodik IV + MH einerseits und allen betrachteten Versionen der Edgington-Methode sowie der RE-Methodik andererseits (siehe x-Achse)

Um verschiedene Fälle, in denen die Edgington-Methode und die kombinierte FE-Methodik IV + MH zu unterschiedlichen Signifikanzaussagen führen, genauer zu beleuchten, sind 4 entsprechende Beispiele in Abbildung 3 dargestellt.

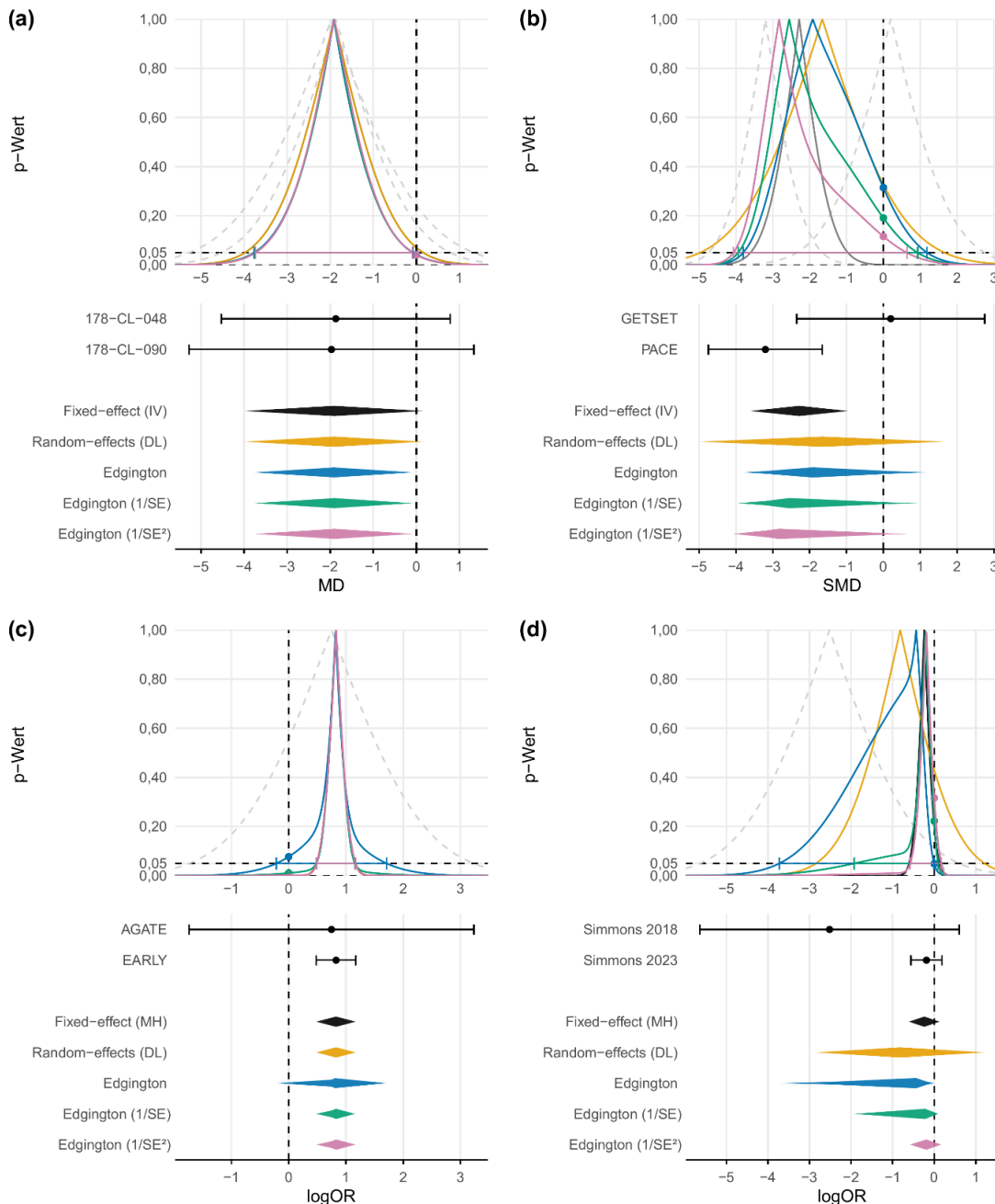


Abbildung 3: Beispiele aus der IQWiG-Metaanalysen-Datenbank, bei denen mindestens 1 Version der Edgington-Methode zu einer anderen Signifikanzaussage als die FE-Methodik IV + MH führt. Unterabbildung (a) zeigt eine Situation mit geringer Diskrepanz zwischen den studienspezifischen Ergebnissen; die Edgington-KIs fallen hier schmäler aus als das FE-KI. Unterabbildung (b) veranschaulicht beispielhaft, wie die verschiedenen Versionen der Edgington-Methode auf Diskrepanz vor allem in den studienspezifischen Punktschätzungen, zum Teil aber auch in den studienspezifischen Standardfehlern, reagieren. Die Unterabbildungen (c) und (d) zeigen beispielhaft, wie die verschiedenen Versionen der Edgington-Methode auf eine stark unterschiedliche Präzision der Studien bei ähnlichen (c) bzw. unterschiedlichen (d) studienspezifischen Punktschätzungen reagieren.

In Abbildung 3(a) sind die Ergebnisse von 2 Studien mit nur geringer Diskrepanz veranschaulicht. In diesem Fall kann das KI der Edgington-Methode schmäler als das der bisher am IQWiG eingesetzten FE-Methodik (hier IV) ausfallen und somit die Signifikanzaussage von nicht statistisch signifikant zu statistisch signifikant ändern. Abbildung 3(b) zeigt 2 Studien, die sehr unterschiedliche Punktschätzungen aufweisen. Hierbei fallen die Edgington-KIs breiter aus als das der FE-Methodik (auch hier IV) und es lässt sich im Gegensatz zu IV kein statistisch signifikanter Gesamteffekt erkennen. Im Beispiel aus Abbildung 3(c) entsteht ein sehr breites KI der ungewichteten Edgington-Methode, was vermutlich auf die stark unterschiedlichen Standardfehler der betrachteten Studien zurückzuführen ist, da die Punktschätzungen der beiden Studien nur geringfügig voneinander abweichen. Im Gegensatz dazu zeigen die gewichteten Versionen der Edgington-Methode sowohl gegenüber der FE- als auch gegenüber der RE-Methodik ein ähnliches Ergebnis. In Abbildung 3(d) sind 2 Studien abgebildet, die sich sowohl in ihren Punktschätzungen als auch in ihren Standardfehlern stark unterscheiden. Hier führt die ungewichtete Edgington-Methode zu einem sehr breiten KI. Die Edgington-Version mit Gewichtung per inversen Varianzen führt zu einem schmalen KI, welches sich stark an der präziseren Studie orientiert, während die Edgington-Version mit Gewichtung per inversen Standardfehlern einen Mittelweg zwischen den beiden anderen Edgington-Versionen darstellt.

6 Diskussion

In diesem Arbeitspapier wurden die Eigenschaften der Edgington-Methode (in der klassischen – ungewichteten – Version sowie in 2 gewichteten Versionen) bei Metaanalysen mit 2 Studien anhand einer empirischen Analyse untersucht. Dazu wurde die Edgington-Methode auf Basis einer Datenbank vergangener IQWiG-Metaanalysen mit den vom IQWiG im Fall von 2 Studien standardmäßig genutzten FE-Verfahren (IV + MH) verglichen. Insgesamt wurden 1226 Metaanalysen mit 2 Studien für die Analyse betrachtet. Die Edgington-Methode (sowohl in der ungewichteten als auch in den 2 gewichteten Versionen) zeigte dabei gute Eigenschaften in Hinblick auf die Breite des zugehörigen KIs sowie in Hinblick auf die Übereinstimmung (mit den bisher am IQWiG genutzten FE-Verfahren) in der Signifikanzaussage des entsprechenden statistischen Tests. Insbesondere sind die Edgington-KIs in allen untersuchten Anwendungen informativ und in der deutlichen Mehrheit der Fälle schmaler als mindestens ein studienspezifisches KI. Eine frühere Untersuchung für Metaanalysen von mehr als 2 Studien [10] zeigte bereits, dass die Breite des metaanalytischen KIs der (ungewichteten) Edgington-Methode je nach Datenlage der zugrunde liegenden Studien variiert: Bei diskrepanten studienspezifischen Punktschätzungen (und teilweise auch bei diskrepanten studienspezifischen Standardfehlern) liefert die Edgington-Methode ein breiteres KI (ohne zur Berechnung dieses KIs explizit einen Heterogenitätsparameter schätzen zu müssen) als in nicht diskrepanten Datensituationen. Die bisher am IQWiG für Metaanalysen mit 2 Studien genutzten FE-Verfahren reagieren nicht in dieser Form auf Diskrepanz in den Studienergebnissen. Empfohlene Methoden für RE-Metaanalysen liefern im Falle von nur 2 Studien häufig extrem breite KIs, was sich insbesondere darin niederschlägt, dass nur sehr selten statistische Signifikanz erreicht wird.

Beispielhaft wurde demonstriert, dass das KI der ungewichteten Edgington-Methode bei hoher Diskrepanz der studienspezifischen Punkt- und / oder Variabilitätsschätzungen sehr breit ausfallen kann. Um die studienspezifischen Variabilitätsschätzungen weiter zu berücksichtigen, lassen sich die gewichteten Versionen der Edgington-Methode verwenden. In den vorgestellten Beispielen war zu erkennen, dass die Gewichtung anhand der inversen Varianz die Punktschätzung stärker in Richtung der Punktschätzung der präziseren Studie verschiebt, während die Gewichtung anhand des inversen Standardfehlers einen Mittelweg zwischen der ungewichteten Edgington-Methode und der Edgington-Methode mit Gewichtung anhand der inversen Varianz darstellt. Daher könnte die Gewichtung anhand des inversen Standardfehlers in Fällen mit sehr unterschiedlichen studienspezifischen Präzisionen zu bevorzugen sein.

Die in diesem Arbeitspapier beschriebene Analyse basiert auf einer empirischen Untersuchung. Sie kann jedoch lediglich eine Voruntersuchung darstellen, da die den Daten zugrunde liegende „Wahrheit“ (d. h. die wahren Werte der Parameter eines angenommenen Modells sowie das angenommene Modell selbst) in einer empirischen Analyse unbekannt ist

und theoretische Eigenschaften der Edgington-Methode (wie Coverage des zugehörigen KIs, Fehler 1. Art des entsprechenden statistischen Tests etc.) somit nicht bewertet werden können. Für eine solche Bewertung und damit letztlich auch für eine Entscheidung für oder gegen den standardmäßigen Einsatz der Edgington-Methode bei Metaanalysen mit 2 Studien in zukünftigen IQWiG-Produkten ist eine Simulationsstudie notwendig. Zwar existiert bereits eine Simulationsstudie für die ungewichtete Edgington-Methode bei Metaanalysen mit 3 Studien (mit guten Ergebnissen für diese Methode) [10], der Fall von 2 Studien ist allerdings noch zu untersuchen. Die Ergebnisse der und Schlussfolgerungen aus der Simulationsstudie werden in einem separaten, noch zu erstellenden Arbeitspapier dargestellt.

7 Fazit

Basierend auf der vorliegenden empirischen Untersuchung der Edgington-Methode für Metaanalysen bei 2 Studien anhand der Datenbank vergangener IQWiG-Metaanalysen könnte die Edgington-Methode eine sinnvolle (möglicherweise sogar besser geeignete) Alternative zu den aktuell am IQWiG im Fall von 2 Studien standardmäßig genutzten FE-Verfahren sein.

Um schlussendlich entscheiden zu können, ob für IQWiG-Metaanalysen mit 2 Studien zukünftig standardmäßig die Edgington-Methode zum Einsatz kommen sollte, bedarf es einer weiteren Untersuchung in Form einer Simulationsstudie. Die Ergebnisse der und Schlussfolgerungen aus der Simulationsstudie werden in einem separaten, noch zu erstellenden Arbeitspapier dargestellt.

8 Literatur

1. Hedges LL, Olkin I. Statistical Methods for Meta-Analysis. Orlando: Academic Press; 1985.
2. Borenstein M, Hedges LV, Higgins JPT et al. Introduction to Meta-Analysis. Chichester: Wiley; 2009.
3. Egger M, Higgins JP, Davey Smith G. Systematic Reviews in Health Research: Meta-Analysis in Context. Hoboken: Wiley; 2022.
4. Bender R, Friede T, Koch A et al. Methods for evidence synthesis in the case of very few studies. Res Synth Methods 2018; 9(3): 382-392. <https://doi.org/10.1002/jrsm.1297>.
5. Guolo A, Varin C. Random-effects meta-analysis: the number of studies matters. Stat Methods Med Res 2017; 26(3): 1500-1518. <https://doi.org/10.1177/0962280215583568>.
6. Schulz A, Schurmann C, Skipka G et al. Performing Meta-analyses with Very Few Studies. Methods Mol Biol 2022; 2345: 91-102. https://doi.org/10.1007/978-1-0716-1566-9_5.
7. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden; Version 8.0 [online]. 2025 [Zugriff: 06.01.2026]. URL: https://doi.org/10.60584/Allgemeine-Methoden_V8.0.
8. Hartung J. An Alternative Method for Meta-Analysis. Biom J 1999; 41(8): 901-916. [https://doi.org/10.1002/\(sici\)1521-4036\(199912\)41:8<901::Aid-bimj901>3.0.Co;2-w](https://doi.org/10.1002/(sici)1521-4036(199912)41:8<901::Aid-bimj901>3.0.Co;2-w).
9. Knapp G, Hartung J. Improved tests for a random effects meta-regression with a single covariate. Stat Med 2003; 22(17): 2693-2710. <https://doi.org/10.1002/sim.1482>.
10. Held L, Hofmann F, Pawel S. A comparison of combined p-value functions for meta-analysis. Res Synth Methods 2025; 16(5): 758-785. <https://doi.org/10.1017/rsm.2025.26>.
11. Pawel S, Roos M, Held L. Combined P-Value Functions for Compatible Effect Estimation and Hypothesis Testing in Drug Regulation. Stat Med 2025; 44(23-24): e70224. <https://doi.org/10.1002/sim.70224>.
12. Bender R, Berg G, Zeeb H. Tutorial: using confidence curves in medical research. Biom J 2005; 47(2): 237-247. <https://doi.org/10.1002/bimj.200410104>.
13. Xie M, Singh K, Strawderman WE. Confidence Distributions and a Unifying Framework for Meta-Analysis. Journal of the American Statistical Association 2011; 106(493): 320-333. <https://doi.org/10.1198/jasa.2011.tm09803>.
14. Irwin JO. On the Frequency Distribution of the Means of Samples from a Population Having Any Law of Frequency with Finite Moments, with Special Reference to Pearson's Type II. Biometrika 1927; 19(3-4): 225-239. <https://doi.org/10.1093/biomet/19.3-4.225>.

15. Hall P. The Distribution of Means for Samples of Size N Drawn from a Population in Which the Variate Takes Values between 0 and 1, All Such Values Being Equally Probable. *Biometrika* 1927; 19(3-4): 240-244. <https://doi.org/10.1093/biomet/19.3-4.240>.
16. Rucker G, Schwarzer G. Beyond the forest plot: The drapery plot. *Res Synth Methods* 2021; 12(1): 13-19. <https://doi.org/10.1002/jrsm.1410>.
17. Xie Mg, Singh K. Confidence Distribution, the Frequentist Distribution Estimator of a Parameter: A Review. *International Statistical Review* 2013; 81(1): 3-39. <https://doi.org/10.1111/insr.12000>.
18. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Wirkungseintritt bei Antibiotika-Therapie; Evidenzbericht zur S3-Leitlinie Therapie entzündlicher Erkrankungen der Gaumenmandeln / Tonsillitis [online]. 2022 [Zugriff: 11.07.2023]. URL: https://www.iqwig.de/download/v21-09b_wirkungseintritt-bei-antibiotika-therapie_evidenzbericht_v1-0.pdf.
19. R Core Team. The R Project for Statistical Computing [online]. [Zugriff: 01.07.2026]. URL: <https://www.R-project.org/>.
20. Fontana S, Hofmann F, Held L et al. confMeta: Confidence Curves and P-Value Functions for Meta-Analysis [online]. 2026 [Zugriff: 26.06.2026]. URL: <https://CRAN.R-project.org/package=confMeta>.
21. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software* 2010; 36(3). <https://doi.org/10.18637/jss.v036.i03>.
22. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer; 2016.
23. Wickham H, Francois R, Henry L et al. dplyr: A Grammar of Data Manipulation [online]. 2026 [Zugriff: 26.06.2026]. URL: <https://CRAN.R-project.org/package=dplyr>.
24. Wickham H, Vaughan D, Girlich M et al. tidyr: Tidy Messy Data [online]. 2025 [Zugriff: 26.06.2026]. URL: <https://CRAN.R-project.org/package=tidyr>.
25. Clarke E, Sherrill-Mix S, Dawson C. ggbeeswarm: Categorical Scatter (Violin Point) Plots [online]. 2025 [Zugriff: 26.06.2026]. URL: <https://CRAN.R-project.org/package=ggbeeswarm>.
26. Kassambara A, Erdey L. ggpubr: 'ggplot2' Based Publication Ready Plots [online]. 2026 [Zugriff: 26.06.2026]. URL: <https://CRAN.R-project.org/package=ggpubr>.
27. Zhu H, Trivison T, Tsai T et al. kableExtra: Construct Complex Table with 'kable' and Pipe Syntax [online]. 2024 [Zugriff: 26.06.2026]. URL: <https://CRAN.R-project.org/package=kableExtra>.
28. Xie Y. Dynamic Documents with R and knitr. Boca Raton: Chapman and Hall/CRC; 2015.

29. Lilienthal J, Sturtz S, Schurmann C et al. Bayesian random-effects meta-analysis with empirical heterogeneity priors for application in health technology assessment with very few studies. Res Synth Methods 2024; 15(2): 275-287. <https://doi.org/10.1002/jrsm.1685>.

30. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. IQWiG-Datenbank für Metaanalysen mit 2 Studien; Stand: 31.12.2025 [online]. [Zugriff: 01.07.2026]. URL: https://www.iqwig.de/download/data_two_studies.xlsx.

Anhang A Offenlegung von Beziehungen der externen Sachverständigen

Im Folgenden sind die Beziehungen der externen Sachverständigen zusammenfassend dargestellt. Alle Informationen beruhen auf Selbstangaben der einzelnen Personen anhand des „Formblatts zur Offenlegung von Beziehungen“. Die in diesem Formblatt aufgeführten Fragen finden sich im Anschluss an diese Zusammenfassung.

Externe Sachverständige

Name	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6	Frage 7
Held, Leonhard	nein	nein	nein	nein	nein	ja	nein
Pawel, Samuel	nein	ja	nein	nein	nein	nein	nein

Im „Formblatt zur Offenlegung von Beziehungen“ wurden folgende 7 Fragen gestellt:

Frage 1: Sind oder waren Sie innerhalb des laufenden Jahres und der 3 Kalenderjahre davor bei einer Einrichtung des Gesundheitswesens (z. B. einer Klinik, einer Einrichtung der Selbstverwaltung, einer Fachgesellschaft, einem Auftragsforschungsinstitut), einem pharmazeutischen Unternehmen, einem Medizinproduktehersteller oder einem industriellen Interessenverband angestellt oder für diese / dieses / diesen selbstständig oder ehrenamtlich tätig bzw. sind oder waren Sie freiberuflich in eigener Praxis tätig?

Frage 2: Beraten Sie oder haben Sie innerhalb des laufenden Jahres und der 3 Kalenderjahre davor eine Einrichtung des Gesundheitswesens (z. B. eine Klinik, eine Einrichtung der Selbstverwaltung, eine Fachgesellschaft, ein Auftragsforschungsinstitut), ein pharmazeutisches Unternehmen, einen Medizinproduktehersteller oder einen industriellen Interessenverband beraten (z. B. als Gutachter/-in, Sachverständige/r, in Zusammenhang mit klinischen Studien als Mitglied eines sogenannten Advisory Boards / eines Data Safety Monitoring Boards [DSMB] oder Steering Committees)?

Frage 3: Haben Sie innerhalb des laufenden Jahres und der 3 Kalenderjahre davor direkt oder indirekt von einer Einrichtung des Gesundheitswesens (z. B. einer Klinik, einer Einrichtung der Selbstverwaltung, einer Fachgesellschaft, einem Auftragsforschungsinstitut), einem pharmazeutischen Unternehmen, einem Medizinproduktehersteller oder einem industriellen Interessenverband Honorare erhalten (z. B. für Vorträge, Schulungstätigkeiten, Stellungnahmen oder Artikel)?

Frage 4: Haben Sie oder hat Ihr Arbeitgeber bzw. Ihre Praxis oder die Institution, für die Sie ehrenamtlich tätig sind, innerhalb des laufenden Jahres und der 3 Kalenderjahre davor von einer Einrichtung des Gesundheitswesens (z. B. einer Klinik, einer Einrichtung der Selbstverwaltung, einer Fachgesellschaft, einem Auftragsforschungsinstitut), einem pharmazeutischen Unternehmen, einem Medizinproduktehersteller oder einem industriellen Interessenverband sogenannte Drittmittel erhalten (d. h. finanzielle Unterstützung z. B. für Forschungsaktivitäten, die Durchführung klinischer Studien, andere wissenschaftliche Leistungen oder Patentanmeldungen)? Sofern Sie in einer größeren Institution tätig sind, genügen Angaben zu Ihrer Arbeitseinheit, z. B. Klinikabteilung, Forschungsgruppe.

Frage 5: Haben Sie oder hat Ihr Arbeitgeber bzw. Ihre Praxis oder die Institution, für die Sie ehrenamtlich tätig sind, innerhalb des laufenden Jahres und der 3 Kalenderjahre davor sonstige finanzielle oder geldwerte Zuwendungen, z. B. Ausrüstung, Personal, Unterstützung bei der Ausrichtung einer Veranstaltung, Übernahme von Reisekosten oder Teilnahmegebühren für Fortbildungen / Kongresse erhalten von einer Einrichtung des Gesundheitswesens (z. B. einer Klinik, einer Einrichtung der Selbstverwaltung, einer Fachgesellschaft, einem Auftragsforschungsinstitut), einem pharmazeutischen Unternehmen,

einem Medizinproduktehersteller oder einem industriellen Interessenverband? Sofern Sie in einer größeren Institution tätig sind, genügen Angaben zu Ihrer Arbeitseinheit, z. B. Klinikabteilung, Forschungsgruppe.

Frage 6: Besitzen Sie Aktien, Optionsscheine oder sonstige Geschäftsanteile einer Einrichtung des Gesundheitswesens (z. B. einer Klinik, einem Auftragsforschungsinstitut), eines pharmazeutischen Unternehmens, eines Medizinprodukteherstellers oder eines industriellen Interessenverbands? Besitzen Sie Anteile eines sogenannten Branchenfonds, der auf pharmazeutische Unternehmen oder Medizinproduktehersteller ausgerichtet ist? Besitzen Sie Patente für ein pharmazeutisches Erzeugnis, ein Medizinprodukt, eine medizinische Methode oder Gebrauchsmuster für ein pharmazeutisches Erzeugnis oder ein Medizinprodukt?

Frage 7: Sind oder waren Sie jemals an der Erstellung einer medizinischen Leitlinie oder klinischen Studie beteiligt, die eine mit diesem Projekt vergleichbare Thematik behandelt/e? Gibt es sonstige Umstände, die aus Sicht von unvoreingenommenen Betrachtenden als Interessenkonflikt bewertet werden können, z. B. Aktivitäten in gesundheitsbezogenen Interessengruppierungen bzw. Selbsthilfegruppen, politische, akademische, wissenschaftliche oder persönliche Interessen?