Universität Ulm
Arbeitsgruppe Klinische Ökonomik

Leitung: Prof. Dr. med. Franz Porzsolt

ICCT – The Internal Consistency of Clinical Trials

Über die inhaltliche Konsistenz randomisierter klinischer Studien

Dissertation
zur Erlangung des Doktorgrades der Medizin
der Medizinischen Fakultät
der Universität Ulm

vorgelegt von

Marc Nagel

geboren in Überlingen

2010

Amtierender Dekan: **Prof. Dr. Thomas Wirth**

1. Berichterstatter: **Prof. Dr. Franz Porzsolt**

2. Berichterstatter: **Frau Prof. Dr. Julia Stingl**

Tag der Promotion: **09.06.2011**

Es gibt auf Erden nur zwei Dinge von Wert:
das erste ist die Liebe,
das zweite - weit dahinter - der Verstand.
Liebe und Verstand sind nie
voneinander geschieden für den,
der ihren Sinn erfasst.

Darüber hinaus gibt es nichts.

Gaston Berger

Inhaltsverzeichnis *Seite*

Abkürzungsverzeichnis **5**

1. Einleitung **6**

2. Methodik **8**

2.1 ICCT – Was ist das überhaupt? 8

2.2 Das Prinzip der vier Ebenen 9

2.3 Die Auswahl der Textstellen 9

2.4 Die Definition wichtiger Begriffe 11

2.5 Die erste Ebene der ICCT – *Die Intervention* 12

2.6 Die zweite Ebene der ICCT – *Das Ergebnis* 13

2.7 Die dritte Ebene der ICCT –*Die durch den Titel geweckten Erwartungen* 15

2.8 Die vierte Ebene der ICCT - *Der primäre Endpunkt im Abstract* 17

2.9 Die Auswahl der wissenschaftlichen Journals 18

2.10 Die Auswahl der Studien 19

2.11 Die Auswertung der ICCT 19

2.12 Die einzelnen Schritte zur Auswertung der Rohdaten 21

3. Ergebnisse **32**

3.1 Das Ergebnis für die fünf Hypothesen 32

3.2 Das Ergebnis zur Identifikation der Problemfelder 34

3.3 Das Ergebnis der Analyse der Handlungsfelder 35

3.4 Das Ergebnis für die Analyse der Endpunkte 39

3.5 Das Ergebnis zur Auswertung der vierten Ebene 45

4. Diskussion **47**

5. Zusammenfassung **61**

6. Literaturverzeichnis **63**

7. Anhang **69**

8. Danksagung **86**

9. Lebenslauf **87**

Abkürzungsverzeichnis

ICCT	=	Internal Consistency of Clinical Trials
EBM	=	Evidence based medicine
CONSORT	=	Consolidated Standards of Reporting Trials

1. Einleitung

Seit dem ausgehenden 19. Jahrhundert begann die zunehmende Professionalisierung der Medizin von der individuellen Erfahrung hin zur wissenschaftlichen Betrachtung von klinischen Fragestellungen. Was zuvor eher im Bereich des Mystischen angesiedelt war, man denke nur an das jahrhundertelange Wirken der Bader und Kurpfuscher, wurde durch so glanzvolle Namen wie Rudolf Virchow oder Robert Koch auf eine zunehmend fundierte wissenschaftliche Basis gestellt. [1] Während die Pioniere der modernen Medizin eher grundsätzliche Fragen beantworteten, deren Tatsachen uns heute unumstößlich erscheinen, entstehen durch den rasanten Fortschritt der heutigen Hochleistungsmedizin neue Fragen und Problemstellungen.

Durch die Vielzahl neuer Therapien und diagnostischer Möglichkeiten tritt zunehmend die Optimierung der medizinischen Versorgung in klinischer und ökonomischer Hinsicht in den Vordergrund des wissenschaftlichen Diskurses. Dies führte zum Ende des letzten Jahrhunderts dazu, dass medizinische Entscheidungen zunehmend auf der Basis der so genannten Evidenz basierten Medizin sprich EBM, getroffen werden. [2]

Das Herzstück dieser EBM sind die randomisierten klinischen Studien. Wichtige klinische Fragestellungen zum Nutzen einer Therapie oder Diagnostik können durch den Vergleich verschiedener Studienpopulationen mit statistisch hinreichender Sicherheit beantwortet werden. [3]

Prinzipiell werden dazu durch die zufällige Verteilung einer definierten Patientenpopulation verschiedene Gruppen erstellt, die in statistischer Hinsicht als annähernd gleich anzusehen sind. [21] Mit einer klinischen Fragestellung, beispielsweise der Wirksamkeit einer Therapie A im Vergleich zu einer Therapie B, wird eine definierte Intervention an diesen Gruppen durchgeführt. Durch den Vergleich der Effekte der beiden Therapien auf das Patientenkollektiv kann nun eine Aussage darüber getroffen werden, welche Therapie den größeren Nutzen darstellt. Dieser Nutzen kann beispielsweise eine Verbesserung der Lebensqualität oder auch der Lebensquantität sein. [5]

Da auf der Grundlage klinischer Studien Entscheidungen getroffen werden, die zum einen eine bedeutende Auswirkung auf die Verteilung zunehmend knapper medizinischer Ressourcen besitzen [13][18] und zum anderen eine direkte Auswirkung auf das individuelle Leben und Überleben des einzelnen Menschen [16], müssen höchste Anforderungen an die Qualität der Studien und die Zuverlässigkeit und Gültigkeiten der Ergebnisse gestellt werden. [1][6]

Dieses Problem der Zuverlässigkeit und Gültigkeit der Ergebnisse klinischer Studien ist ein intensiv diskutierter Gegenstand der aktuellen wissenschaftlichen Diskussion. [6] Bereits 1980 stelle *Hemminki* im British Medical Journal fest, dass viele klinische Studien von schlechter methodischer Qualität sind da wichtige Informationen, beispielsweise zu unerwünschten Nebenwirkungen einer Therapie, nur unzureichend dargestellt werden. [57] Zur Verbesserung der Qualität klinischer Studien wurden seither wichtige Schritte unternommen. So publizierten *Moher et al* 2001 eine revidierte Version des bereits durch *Begg et al* 1996 vorgestellten CONSORT Statement [4][67] zur Verbesserung der Qualität der Berichterstattung randomisierter klinischer Studien. Es wurde eine Checkliste mit 22 Punkten gefordert, die in jeder klinischen Studie vorhanden sein sollten. Beispielsweise sollten die Gruppenspezifikationen, die definierten Endpunkte oder die Art der Verblindung genau beschrieben werden. [4][33] *Krleza-Jeric et al* publizierten 2005 das Ottawa Statement in dem eine Registrierung und vollständig Veröffentlichung des Studienprotokolls jeder klinischen Studie gefordert wurde. [39] Auch von Seiten der wissenschaftlichen Journals wurden Schritte unternommen, die zu einer Verbesserung der Qualität klinischer Studien führen sollten. Beispielsweise durch eine strengere Revision der eingereichten Arbeiten. [8] Diese Maßnahmen führten insgesamt zu einer Verbesserung der Berichterstattung klinischer Studien. [9]

Dennoch existieren weiterhin Mängel, die Raum für Verbesserungen bieten. So ist es ein häufiges Phänomen, dass relevante Informationen über die Ergebnisse klinischer Studien zurückgehalten werden. Eine klinische Entscheidungsfindung auf Grundlage dieser Ergebnisse wird dadurch erschwert. [17] Ebenso wird die statistisch notwendige Größe der Studiengruppen oft unterschritten. Das führt dazu, dass die Beweiskraft einer klinischen Studie zu gering wird, um einen signifikanten Unterschied zwischen den Gruppen zu finden. [7][14][15] Häufig bestehen inhaltliche Unterschiede zwischen den eingereichten Studienprotokollen und den endgültig publizierten Arbeiten. [11] Ein großes Problemfeld ist die Auswertung der Daten klinischer Studien. So kann man je nach Art der Auswertung zu unterschiedlichen Ergebnissen kommen. Hier sei nur der „Intention to treat“ Ansatz in der Datenauswertung im Vergleich zur Auswertung der Daten „per protocol“ erwähnt. [31] Viele Studien definieren erst nach der Datensammlung Subgruppen. Für diese wird ein klinischer Effekt beobachtet, der durch die Datenauswertung dem ursprünglichen Studienprotokoll entsprechend nicht beobachtet worden wäre. Das erhöht das Risiko falsch positiver Ergebnisse. [51] Die Darstellung der Ergebnisse im Abstract klinischer Studien gibt oft ein

verzerrtes Bild der wahren Ergebnisse wieder, mit dem Resultat dass klinische Effekte oft überschätzt werden. [47]

Chan et al publizierten 2004 eine Arbeit im *Journal of American Medial Association* in der sie feststellten, dass Studienergebnisse häufig nicht nur unvollständig, sondern im Vergleich mit den vorliegenden Studienprotokollen auch häufig inkonsistent und fehlerhaft sind.[12] Die hier vorliegende Arbeit ICCT –„The Internal Consistency of Clinical Trials“ „Über die inhaltliche Konsistenz randomisierter klinischer Studien“ möchte an diesem Punkt anknüpfen. Da die Protokolle vieler klinischer Studien im Inhalt nicht konsistent mit den Originalarbeiten sind, möchte die ICCT herausfinden wie konsistent dann der Inhalt der Originalarbeiten ist. Dazu wird diese Arbeit definierte Textstellen innerhalb einer klinischen Studie miteinander vergleichen und überprüfen ob der Inhalt dieser Textstellen übereinstimmt oder nicht. In den Worten von *Gravel et al* stellt die ICCT folgende Frage: *Are authors saying what they do and doing what they say?* [31]

2. Methodik

2.1 ICCT – Was ist das überhaupt?

Die ICCT bewertet die inhaltliche Konsistenz einer randomisierten klinischen Studie. Während sich viele bisherigen Arbeiten der Qualität klinischer Studien angenähert haben, in dem sie beispielsweise die Studienprotokolle mit den publizierten Arbeiten verglichen [12] oder die Qualität der Verblindung innerhalb von Studien untersuchten, [23] verfolgt diese Arbeit einen neuen Ansatz. Die ICCT hat durch das einfache Lesen einer Studie den Inhalt von Textstellen miteinander verglichen und geprüft, ob dieser Inhalt konsistent ist. Dazu wurden verschiedene Textstellen innerhalb der Studie definiert. In diesen Textstellen wurden bestimmte Schlüsselpunkt aufgesucht und deren Inhalt durch reines Gegenlesen miteinander verglichen. Beispielsweise hat die ICCT die Endpunkte einer klinischen Studie herausgesucht und geprüft, ob für diese Endpunke auch Ergebnisse präsentiert wurden.

Um die inhaltliche Konsistenz einer randomisierten klinischen Studie beurteilen zu können wurde eine Methode entwickelt, mit deren Hilfe eine objektive Antwort auf die Fragestellung nach der Konsistenz dieser klinischen Studie gegeben werden kann. Diese Methode wird nun in den nachfolgenden Kapiteln ausführlich vorgestellt.

2.2 Das Prinzip der vier Ebenen

Jede klinische Studie ist nach dem gleichen Prinzip aufgebaut. Die Einleitung stellt die Arbeit kurz vor. Es wird dargestellt, welche Untersuchung durchgeführt werden soll und aus welchem Grund. [20] In der Methodik wird die genaue Herangehensweise an die klinische Fragestellung vorgestellt. Dazu gehören die Definition der Intervention, der Studiengruppe und der Endpunkte. Im Ergebnisteil werden die gefundenen Resultate wertneutral vorgestellt. Den Schluss jeder Studie bildet die Diskussion, sozusagen das Herzstück der Studie. Hier werden die gefundenen Ergebnisse bewertet und es werden Schlussfolgerungen gezogen, die für das weitere klinische Vorgehen wichtig sind. Beispielsweise wird die Schlussfolgerung gezogen, dass Therapie A besser wirkt als Therapie B. [22]

Die ICCT stellte vier Fragen an jede Studie:

1. Welche Intervention war in der klinischen Studie geplant?
2. Welche Intervention wurde durchgeführt?
3. Welche Endpunkte wurden definiert?
4. Welche Ergebnisse wurden gefunden?

Um diese Fragen objektiv beantworten zu können wurde durch einen Kunstgriff jede Studie in vier Ebenen unterteilt. Jede Ebene vergleicht dabei den Inhalt bestimmter Stellen im Text. Um den Inhalt dieser Textstellen vergleichen zu können, definiert die ICCT Schlüsselpunkte für jede Textstelle. Alle Informationen zu den Schlüsselpunkten werden gesammelt. Anschließend werden diese Informationen verglichen. Dadurch kann eine Aussage darüber getroffen werden, ob die Textstellen in ihrem Inhalt konsistent sind.

2.3 Die Auswahl der Textstellen

Jede Studie kann in sechs verschiedene Abschnitte unterteilt werden. Der erste Abschnitt ist der Titel der Studie. Der zweite Abschnitt ist die Einleitung, gefolgt vom dritten Abschnitt, der Methodik. Den vierten Abschnitt bildet der Ergebnisteil. Im fünften Abschnitt folgt die Diskussion. Den Abschluss bildet als sechster Abschnitt der Abstract. [17][20][22] Die vier Ebenen der ICCT vergleichen den Inhalt von fünf dieser sechs Textstellen.

Die erste Textstelle: *Der Titel*

Der Titel einer Studie gibt eine erste Auskunft darüber, welche Intervention in dieser Studie durchgeführt werden wird und welche Fragestellung beantwortet werden soll.

Die zweite Textstelle: *Die Einleitung*

Die Einleitung erläutert in kurzen Worten den Grund für die Durchführung der Studie. Am Ende der Einleitung wird meistens eine Zielgruppe für die Studie genannt. Die Intervention der Studie wird kurz vorgestellt und es wird eine konkrete Fragestellung an die Studie formuliert.

Die dritte Textstelle: *Die Methodik*

Die Methodik beschreibt ausführlich das Studiendesign. Die Studiengruppe mit ihren spezifischen Charakteristika und die genaue Intervention werden vorstellt. Die Endpunkte werden definiert und deren Erhebung erklärt.

Die vierte Textstelle: *Die Diskussion*

Die Diskussion ist der wichtigste Teil jeder Studie. Sie stellt die Ergebnisse der Studie in einen wissenschaftlichen Kontext und zieht entsprechende Schlussfolgerungen daraus. Im Gegensatz zum Ergebnisteil, der die Ergebnisse ohne Wertung vorstellt.

Die fünfte Textstelle: *Der Abstract*

Der Abstract fasst alle Informationen aus einer Studie kurz zusammen. Er beschreibt die Studiengruppe, die Intervention und die Endpunkte. Auch die gefundenen Ergebnisse für die Endpunkte stehen im Abstract. Das Herzstück im Abstract bildet die *Schlussfolgerung.* Sie fasst die wichtigsten Aussagen der Studie kurz zusammen.

Der Ergebnisteil wurde in der ICCT nicht als Textstelle benutzt, da die Ergebnisse dort nur präsentiert werden ohne in ihrer klinischen Bedeutung bewertet zu werden. Als Textstelle für die beobachten Ergebnisse einer klinischen Studie wurde die Diskussion gewählt. Sie stellt die Ergebnisse in einen klinischen Kontext und zieht daraus Schlussfolgerungen für die Anwendung dieser Ergebnisse in der Praxis.

2.4 Definition wichtiger Begriffe

Bevor die vier Ebenen der ICCT im Detail vorgestellt werden, müssen noch einige Begriffe definiert werden, die für das Verständnis dieser Arbeit wichtig sind.

Intervention:
Die ICCT definiert die Intervention in einer Studie als die Behandlung oder das Untersuchungsverfahren dem die Studiengruppen durch Zufallsverteilung zugeordnet werden.

Die erwartete Intervention:
Die ICCT definiert als die erwartete Intervention die Behandlung oder das Untersuchungsverfahren, das die zu untersuchende Studie entweder im Titel oder in der Einleitung vorstellt.

Die durchgeführte Intervention:
Als durchgeführte Intervention definiert die ICCT die Behandlung oder das Untersuchungsverfahren, das in der zu untersuchenden Studie im Methodikteil vorgestellt wird.

Das erwartete Ergebnis:
Die Methodik definiert Endpunkte um eine Antwort auf eine bestimmte Fragestellung zu erhalten. Das erwartete Ergebnis entspricht also der Fragestellung aus den definierten Endpunkten der Methodik.

Das beobachtete Ergebnis:
Die ICCT definiert als beobachtetes Ergebnis alle Ergebnis aus der Diskussion und alle Ergebnisse aus dem Abstract.

- Die *erwartete Intervention* entspricht der Intervention im **Titel** und in der **Einleitung**
- Die *durchgeführte Intervention* ist die Intervention in der **Methodik**
- Das *erwartete Ergebnis* sind die definierten Endpunkte der **Methodik**
- Das *beobachtete Ergebnis* sind die Ergebnisse der **Diskussion** uns des **Abstract**

2.5 Die erste Ebene der ICCT – Die Intervention

In der ersten Ebene vergleicht die ICCT die Textstellen ***Einleitung*** und ***Methodik***

Die erste Ebene prüft die inhaltliche Konsistenz der Einleitung im Vergleich mit der Methodik. Dabei soll die erwartete Intervention der Studie mit der durchgeführten Intervention der Studie verglichen werden. Jeder Intervention besteht aus drei Teilen.

- Der Studiengruppe
- Der Intervention an dieser Studiengruppe
- Der Fragestellung die beantwortet werden soll

Jeder dieser drei Teile entspricht einem Schlüsselpunkt der ersten Ebene. Es werden alle Informationen zu diesen drei Schlüsselpunkten an beiden Textstellen gesammelt. Sind die Informationen identisch, so stimmt der Schlüsselpunkt überein.

Der erste Schlüsselpunk DIE STUDIENGRUPPE

Hier vergleicht man die Studiengruppe, die in der Einleitung kurz beschrieben ist, mit der Studiengruppe, die in der Methodik definiert wird. Dabei wird geprüft, ob diese Gruppen identisch sind, oder ob sie sich in wichtigen Merkmalen unterscheiden.

Der zweite Schlüsselpunkt DIE INTERVENTION

Der zweite Schlüsselpunkt vergleicht die Behandlung oder das Untersuchungsverfahren, das in der Einleitung kurz beschrieben wird mit der Behandlung oder dem Untersuchungsverfahren aus der Methodik.

Der dritte Schlüsselpunkt DAS ZIEL DER INTERVENTION

Die Einleitung einer Studie stellt meistens konkrete Fragestellungen an diese Studie. Im dritten Schlüsselpunkt wird überprüft, ob diese Fragestellungen aus der Einleitung mit den definierten Endpunkten aus der Methodik übereinstimmen.

Tabelle 1: Die Schlüsselpunkte der ersten Ebene

Die erste Ebene - Die Schlüsselpunkte-	Textstelle *Einleitung*	Textstelle *Methodik*	*Übereinstimmung der Informationen?*
Die Studiengruppe			**ja/nein/möglich**
Die Intervention			**ja/nein/möglich**
Das Ziel der Intervention			**ja/nein/möglich**

Aus Gründen der Übersichtlichkeit finden sich die genauen Definitionen zum Auffinden der Schlüsselpunkte in den Textstellen im Anhang der Arbeit. Das Gleiche gilt für die Definitionen zur inhaltlichen Übereinstimmung der Schlüsselpunkte.

2.6 Die zweite Ebene der ICCT – Das Ergebnis

In der zweiten Ebene vergleicht die ICCT die Textstellen ***Diskussion***, ***Methodik*** und ***Abstract***

Das Prinzip der randomisierten klinischen Studien wurde entwickelt, um eine Antwort auf medizinisch wichtige Fragen zu erhalten. [1] In klinischen Studien stellt man Fragen, indem man Endpunkte definiert. [22] Für diese Endpunkte erhält man am Ende ein Ergebnis. Interessiert zum Beispiel das Langzeitüberleben durch eine bestimmte Therapie, dann definiert man als Endpunkt die Zahl der Überlebenden jeder Studiengruppe. Überleben mehr Patienten in der Gruppe A als in der Gruppe B, dann kann man daraus schließen, dass Therapie A das Überleben wahrscheinlich erhöht hat.
Definierte Endpunkte besitzen eine unterschiedliche Wertigkeit. Der wichtigste Endpunkt einer Studie ist der primäre Endpunkt. Er ist die eigentliche Fragestellung der jeweiligen Studie. Das ist wichtig zu wissen, da die Zahl der benötigten Studienteilnehmer anhand des primären Endpunktes berechnet werden muss. [14] Sekundäre und zusätzliche Endpunkte können helfen weitere Aspekte und zusätzliche Fragen zu klären. Dabei stehen deren Ergebnisse in der Wertigkeit jedoch hinter dem Ergebnis für den primären Endpunkt. [50]

Die ICCT vergleicht in der zweiten Ebene das erwartete Ergebnis mit dem beobachteten Ergebnis. Dazu werden in dieser Ebene die Textstellen Methodik, Diskussion und Abstract verglichen. Es werden alle Informationen zu den Schlüsselpunkten dieser Textstellen aufgesucht und dann verglichen. Sind die Informationen gleich, so stimmen die Schlüsselpunkte überein.

Aus Gründen der Übersichtlichkeit wurde die zweite Ebene nochmals unterteilt. In der zweiten Ebene E2A werden die Schlüsselpunkte an den Textstellen Methodik und Diskussion verglichen. In der zweiten Ebene E2B vergleicht die ICCT dieselben Schlüsselpunkte innerhalb der Diskussion und des Abstract.

Der erste Schlüsselpunkt DIE STUDIENGRUPPE

Die ICCT vergleicht in der zweiten Ebene E2A als ersten Schlüsselpunkt die Studiengruppe aus der Methodik mit der Studiengruppe der Diskussion. Dabei wird geprüft, ob die Informationen zu diesen beiden Gruppen identisch sind oder ob sie sich in wichtigen Merkmalen unterscheiden.

In der zweiten Ebenen E2B wird als erster Schlüsselpunkt die Studiengruppe der Diskussion mit der Studiengruppe des Abstracts verglichen.

Der zweite Schlüsselpunkt DAS ERGEBNIS

Dieser zweite Schlüsselpunkt ist der wichtigste Teil der ICCT. Er vergleicht in der zweiten Ebene E2A die definierten Endpunkte aus der Methodik mit den beobachteten Ergebnissen in der Diskussion. Die ICCT wird prüfen, ob für jeden definierten Endpunkt ein Ergebnis beobachtet wird. Sie wird auch prüfen ob für jedes beobachtete Ergebnis ein Endpunkt definiert war.

In der zweiten Ebene E2B vergleicht dieser Schlüsselpunkt die beobachteten Ergebnisse der Diskussion mit den Ergebnissen im Abstract.

Tabelle 2: Die Schlüsselpunkte der zweiten Ebene E2A

Die zweite Ebene E2A	Textstelle *Methodik*		Textstelle *Diskussion*	*Übereinstimmung der Informationen?*
Studiengruppe		*Studiengruppe*		**ja/nein/möglich**
Definierte Endpunkte		*Beobachtete Ergebnisse*		**ja/nein/möglich**

Tabelle 3: Die Schlüsselpunkte der zweiten Ebene E2B

Die zweite Ebene E2B	Textstelle *Diskussion*		Textstelle *Abstract*	*Übereinstimmung der Informationen?*
Studiengruppe		*Studiengruppe*		**ja/nein/möglich**
beobachtete Ergebnisse		*Beobachtete Ergebnisse*		**ja/nein/möglich**

Aus Gründen der Übersichtlichkeit finden sich die genauen Definitionen zum Auffinden der Schlüsselpunkte in den Textstellen auch für die zweite Ebene im Anhang der Arbeit. Das Gleiche gilt für die Definitionen zur inhaltlichen Übereinstimmung der Schlüsselpunkte.

Der Unterschied zwischen den Ebenen E2A und E2B

In der zweiten Ebene vergleicht die ICCT das erwartete Ergebnis mit dem beobachteten Ergebnis. Wie aus den Definitionen folgt, entspricht das erwartete Ergebnis den definierten Endpunkten der Methodik. Das beobachtete Ergebnis entspricht den Ergebnissen in der Diskussion. Im Abstract und in der Diskussion finden sich definitionsgemäß nur die beobachteten Ergebnisse.

Um also die beobachteten Ergebnisse mit den erwarteten Ergebnissen vergleichen zu können, wurde die Ebene E2A eingeführt. Sie vergleicht die beiden Schlüsselpunkte zwischen den Textstellen Methodik und Diskussion.

Um die inhaltliche Konsistenz des beobachteten Ergebnisses selbst zu prüfen, wurde die Ebene E2B eingeführt, da sich das erwartete Ergebnis definitionsgemäß an zwei Textstellen befindet, dem Abstract und der Diskussion.

2.7 Die dritte Ebene der ICCT –Die durch den Titel geweckten Erwartungen

In der dritten Ebene vergleicht die ICCT die Textstellen ***Titel***, ***Methodik*** und ***Diskussion***

Die Erwartung [des Lesers] *wird zunächst durch den Titel geweckt, der andeuten und Grundspannung schaffen soll*. [68] Dieses aus der Literaturwissenschaft bekannte Prinzip der Erzeugung einer Erwartungshaltung an einen literarischen Text kann auch auf die Betitelung wissenschaftlicher Arbeiten übertragen werden.

Die Betitelung einer wissenschaftlichen Arbeit dient dem Leser als Blickfang. Durch die dort gegebenen Informationen fällt eine erste Entscheidung, ob diese Arbeit für die eigenen Belange von Interesse sein könnte oder nicht. Der Titel einer klinischen Studie enthält meistens eine kurze Beschreibung der Fragestellung, die durch eine Intervention geklärt werden soll.

In der dritten Ebene vergleicht die ICCT die Textstellen Titel, Methodik und Diskussion. Es soll geprüft werden, ob die durch den Titel geweckten Erwartungen an die Intervention und die Fragestellungen der Studie tatsächlich erfüllt werden. An den drei genannten Textstellen vergleicht die ICCT zwei Schlüsselpunkt.

Der erste Schlüsselpunkt DIE INTERVENTION

Die meisten Studien beschreiben prinzipiell bereits im Titel, welche Intervention diese Studie durchführen wird. Die ICCT vergleicht, ob die im Titel beschriebene Intervention mit der durchgeführten Intervention aus der Methodik übereinstimmt.

Der zweite Schlüsselpunkt DAS ERGEBNIS

Die meisten Titel einer Studie formulieren bereits eine konkrete Fragestellung. Der zweite Schlüsselpunktes prüft, ob diese konkrete Fragestellung des Titels in der Diskussion durch ein beobachtetes Ergebnis beantwortet wird.

Aus praktischen Gründen wird auch die dritte Ebene geteilt. Die Ebene E3A vergleicht den ersten Schlüsselpunkt in den Textstellen Titel und Methodik. Die Ebene E3B vergleicht den zweiten Schlüsselpunkt in den Textstellen Titel und Diskussion.

Tabelle 4: Der Schlüsselpunkt der dritten Ebene E3A

Die dritte Ebene E3A	Textstelle *Titel*		Textstelle *Methodik*	*Übereinstimmung der Informationen?*
Die erwartete Intervention		*Die durchgeführte Intervention*		**ja/nein/möglich**

Tabelle 5: Der Schlüsselpunkt der dritten Ebene E3B

Die dritte Ebene E3B	Textstelle *Titel*		Textstelle *Diskussion*	*Übereinstimmung der Informationen?*
Das erwartete Ergebnis		*Das beobachtete Ergebnis*		**ja/nein/möglich**

Aus Gründen der Übersichtlichkeit finden sich die genauen Definitionen zum Auffinden der Schlüsselpunkte in den Textstellen auch für die dritte Ebene im Anhang der Arbeit. Das Gleiche gilt für die Definitionen zur inhaltlichen Übereinstimmung der Schlüsselpunkte.

Der Unterschied zwischen den Ebenen E3A und E3B

Prinzipiell ist die dritte Ebene „Die durch den Titel geweckten Erwartungen“ eine Mischung aus der ersten Ebene „Die Intervention“ und der zweiten Ebene „Das Ergebnis“. Es wird in Kurzform geprüft ob die beiden Schlüsselpunkte *DIE INTERVENTION* und *DAS ERGEBNIS* innerhalb der Textstellen Titel, Methodik und Diskussion inhaltlich übereinstimmen.

Da die beiden Schlüsselpunkte unterschiedlichen Textstellen zugeordnet sind, wurde die dritte Ebene aus Gründen der Übersichtlichkeit getrennt. Die Ebene E3A vergleicht den Titel mit der Methodik und prüft, ob die im Titel beschriebene Intervention der Intervention der Methodik entspricht. Die Ebene E3B vergleicht den Titel mit der Diskussion und prüft, ob die Fragestellung im Titel durch ein beobachtetes Ergebnis in der Diskussion beantwortet wird.

2.8 Die vierte Ebene der ICCT - Der primäre Endpunkt im Abstract

In der vierten Ebene vergleicht die ICCT die ***Schlussfolgerung*** im Abstract mit dem primären Endpunkt der ***Methodik***

Der Abstract einer klinischen Studie fasst die Ergebnisse der Studie in Kurzform zusammen. Dass der Abstract klinischer Studien in vielen Fällen nicht alle Ergebnisse der Studie enthält ist ein bekanntes Problem. [43][44] [45] Dadurch besteht die Gefahr, dass aus den Ergebnissen des Abstracts falsche Schlussfolgerungen gezogen werden. Das kann dazu führen, dass die klinischen Effekte einer Therapie überschätzt werden. [46][47]
Die vierte Ebene unterscheidet sich von bisherigen drei Ebenen. Sie untersucht nicht komplette Textstellen, sondern geht nur einer einzigen Fragestellung nach:

Bezieht sich die Schlussfolgerung im Abstract auf das beobachtete Ergebnis für den primären Endpunkt?

Die Power einer Studie gibt an, mit welcher Wahrscheinlichkeit ein statistisch signifikanter Unterschied zwischen den Studiengruppen auch in Wahrheit besteht. Die Power hängt von der Größe der Studiengruppe und dem Endpunkt ab. [24][25] Jede klinische Studie stellt zudem Hypothesen auf. Zur Prüfung dieser Hypothesen werden passende Endpunkte definiert. Über die Gültigkeit einer Hypothese kann aber nur dann eine Aussage gemacht werden, wenn alle statistischen Größen passend für den entsprechenden Endpunkt berechnet wurden. [26][27][28].
Der primäre Endpunkt ist das Herzstück der Studie. Auf seiner Basis werden alle wichtigen Größen der Studie berechnet. Also ist er die einzige Fragestellung einer klinischen Studie, die mit statistisch hinreichender Wahrscheinlichkeit beantwortet werden kann. Folglich sollte sich die Schlussfolgerung im Abstract auf das Ergebnis für den primären Endpunktes beziehen.

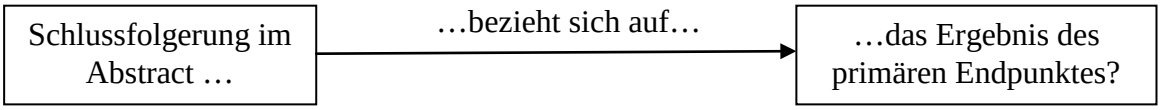

Abbildung 1: Die vierte Ebene – Bezugnahme der Conclusion im Abstract auf den primären Endpunkt

Aus Gründen der Übersichtlichkeit werden auch die Definitionen für die Auswertung der Daten in der vierten Ebene im Anhang dieser Arbeit vorgestellt.

2.9 Die Auswahl der wissenschaftlichen Journals

Gluud et al stellten 2005 in einem Artikel für das *American Journal of Gastroenterology* fest, dass die Qualität klinischer Studien umso besser sei, je höher der Impact Faktor des publizierenden Journals ist. [19] Der Impact Faktor eines Journals gibt an, mit welcher durchschnittlichen Häufigkeit eine Arbeit in diesem Journal durch andere Arbeiten zitiert werden wird.
Um also qualitativ möglichst hochwertige klinische Studien durch die ICCT auf ihre inhaltliche Konsistenz zu überprüfen, wurden die Studien aus Journals herausgesucht, die einen möglichst hohen Impact Faktor aufweisen. Als Grundlage für den Impact Faktor diente der *Journal Citation Report* von 2003 [72].
Trotz eines hohen Impact Faktors konnten nicht alle geeigneten Journals für diese Arbeit berücksichtigt werden, da der Volltextzugriff im Netzwerk der Universität Ulm nur auf bestimmte Journals möglich ist. Der individuelle Zugriff auf die Volltexte anderer Journals hätte einen unverhältnismäßig großen finanziellen Aufwand bedeutet, sodass diese Journals nicht berücksichtigte werden konnten.
Die klinischen Studien für die ICCT wurden aus folgenden Journals herausgesucht. Der Impact Faktor des Jahres 2003 ist in Klammern angefügt. [72]

- *New England Journal of Medicine (34,833)*
- *American Journal of Obstetrics and Gynecology (02,518)*
- *The Lancet (18,316)*

Wegen fehlendem Volltextzugriff konnten diese Journals nicht genutzt werden:
British Medical Journal, Journal of American Medical Association und *Surgery*

2.10 Auswahl der Studien

Die Internetauftritte der gewählten Journals wurden nach folgenden Schlüsselwörtern durchsucht: *Randomized*, *Controlled* und *Trials*. [69][70][71] Die Themen der Studien spielten keine Rolle für die Auswahl, jede bis August 2005 publizierte randomisierte klinische Studie wurde gewählt bis von allen drei Internetauftritten insgesamt 53 Studien im Volltext herunter geladen worden waren. Dabei wurde die Anzahl der Studien ungefähr gleichmäßig auf die Journals verteilt. Die in dieser Arbeit bewerteten Studien sind im Literaturverzeichnis angegeben. [79] bis [131]

2.11 Die Auswertung der ICCT

Die ICCT wurde entwickelt, um die inhaltliche Konsistenz randomisierter klinischer Studien zu prüfen. Dazu wurden verschiedene Schlüsselpunkte definiert und deren inhaltliche Konsistenz an verschiedenen Textstellen verglichen.

Die Auswertung der ermittelten Daten für diese Schlüsselpunkte erfolgt dabei rein deskriptiv. Es wird die Anzahl der Studien gezählt, die inhaltlich in einem Schlüsselpunkt übereinstimmen. Ebenfalls wird die Anzahl der Studien gezählt, die an dem betreffenden Schlüsselpunkt nicht übereinstimmen. Dabei ist eine Entscheidung, ob eine Übereinstimmung besteht oder nicht, nicht immer möglich. Also wird auch die Anzahl derjenigen Studien angegeben, die an einem bestimmten Schlüsselpunkt möglicherweise übereinstimmen.

Insgesamt werden für diese Arbeit 53 Studien aus 3 Journals bearbeitet. Neben der absoluten Anzahl der inhaltlichen Übereinstimmungen, Nichtübereinstimmungen oder möglichen Übereinstimmungen an einem Schlüsselpunkt wird diese Anzahl auch in Prozent aller bearbeiteten Studien angegeben werden.

Durch die Unterteilung jeder Studie in vier verschiedene Ebenen ist es möglich, aus den gefundenen Daten weitere Schlüsse zu ziehen. Beispielsweise kann festgestellt werden, welche Ebene am häufigsten inhaltlich nicht übereingestimmt hat. Die gleiche Aussage lässt sich für die einzelnen Schlüsselpunkte in dieser Ebene treffen. Die Rohdaten der ICCT werden nach einem festen Schema ausgewertet. Dabei werden verschiedene Auswertungsschritte aufeinander folgen, die eine immer genauere Analyse der gefundenen Daten erlauben.

Im ersten Schritt werden fünf Hypothesen erstellt und geprüft. Anschließend wird die inhaltliche Konsistenz der ersten drei Ebenen ermittelt. Im nächsten Schritt ist dann eine genaue Analyse der inhaltlichen Konsistenz der einzelnen Schlüsselpunkte geplant. Zuletzt können durch die erhobenen Rohdaten der ICCT zusätzliche Analysen durchgeführt werden Zuletzt erfolgt die Auswertung der vierten Ebene. Die einzelnen Auswertungsschritte werden nun im Folgenden ausführlich beschrieben.

Die Sammlung der Rohdaten – Der Auswertungsbogen

Um eine standardisierte Bewertung der inhaltlichen Konsistenz einer randomisierten klinischen Studie durchzuführen, wurde für diese Arbeit ein eigener Auswertungsbogen erstellt. Er findet sich im Anhang dieser Arbeit.
Alle Informationen zu den definierten Schlüsselpunkten wurden getrennt für jede der vier Ebenen in den Textstellen gesammelt und dann von Hand in den Auswertungsbogen übertragen. Anschließend wurde jeder Schlüsselpunkt der einen Textstelle mit dem Schlüsselpunkt der zu vergleichenden Textstelle auf dem Bogen gegengelesen und anhand von definierten Kriterien die Entscheidung getroffen, ob diese beiden Schlüsselpunkte inhaltlich übereinstimmen oder nicht. War eine Entscheidung nicht möglich so wurde festgelegt, dass die beiden Schlüsselpunkte möglicherweise übereinstimmen. Die definierten Kriterien finden sich, nach den jeweiligen Ebenen getrennt, im Anhang dieser Arbeit.
Die vierte Ebene – Der primäre Endpunkt im Abstract- besitzt keine eigene Spalte auf dem Auswertungsbogen. Bei der Erstellung des Auswertungsbogens war die vierte Ebene noch kein endgültiger Bestandteil der ICCT. Es wurde jedoch handschriftlich auf der Rückseite jedes Auswertungsbogens vermerkt, ob sich die Schlussfolgerung im Abstract auf das beobachtete Ergebnis für den primären Endpunkt bezieht.
Wenn am Schlüsselpunkt *DAS ERGEBNIS* in der Textstelle Methodik ein Endpunkt gefunden wurde, ist auf dem Auswertungsbogen handschriftlich vermerkt, ob es sich dabei um den primären Endpunkt, einen sekundären Endpunkt oder einen zusätzlichen Endpunkt handelt.
Reichte der Platz in den entsprechenden Feldern des Auswertungsbogens nicht für alle Informationen über die Schlüsselpunkte, so wurde entweder ein Extrablatt hinzugenommen oder die unbedruckte Rückseite des Bogens beschrieben. In den

entsprechenden Feldern des Auswertungsbogens wurde ein Vermerk eingetragen wo die weiteren Informationen zum jeweiligen Schlüsselpunkte zu finden sind.

2.12 Die einzelnen Schritte zur Auswertung der Rohdaten

Die Auswertung der ICCT erfolgt in vier Schritten. Dabei ermöglicht jeder Schritt eine detailliertere Aussage über die Ursachen einer fehlenden inhaltlichen Konsistenz der Studie. Die Auswertungsschritte sollen es am Ende ermöglichen, auf Grundlage der gefundenen Ergebnisse dieser Arbeit, entsprechende Schlussfolgerungen zu ziehen, die eine Einordnung in die bisherige wissenschaftliche Diskussion erlauben und Möglichkeiten aufzuzeigen, wie die Berichterstattung klinischer Studien verbessert werden kann.

Graphisch gesehen entsprechen die vier aufeinander folgenden Auswertungsschritte einer Pyramide. Von der Basis bis zur Spitze der Pyramide ermöglicht diese Vorgehensweise eine immer detailreichere Analyse der Rohdaten.

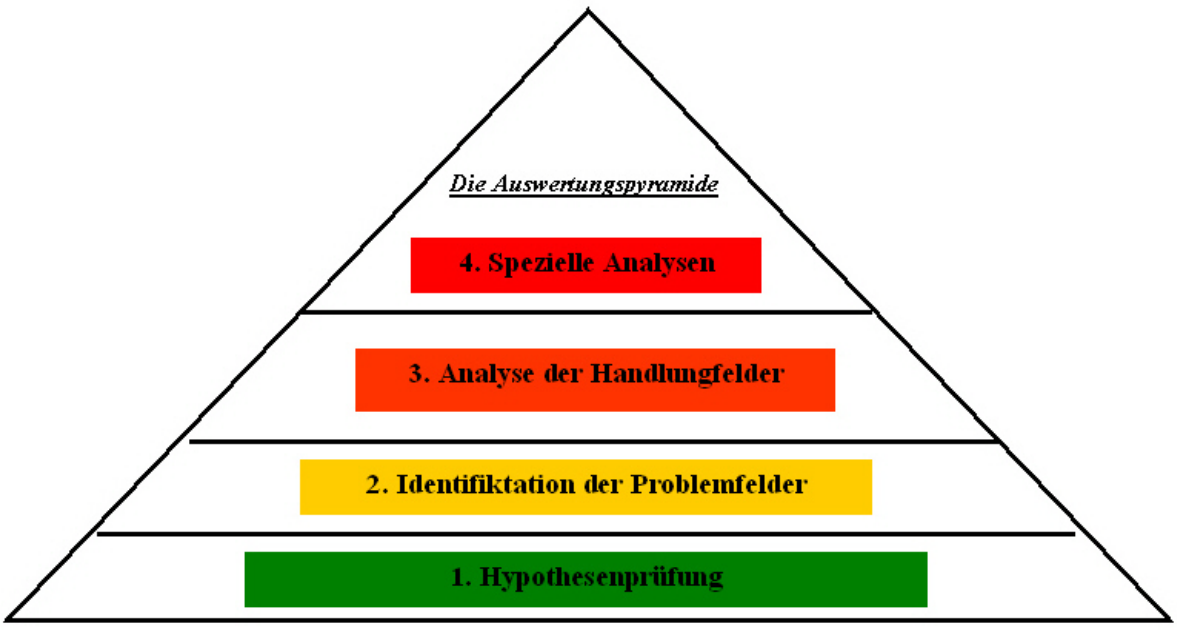

Abbildung 2: Die Auswertungspyramide

Der erste Auswertungsschritt: Die Prüfung der fünf Hypothesen

Die ICCT erstellt fünf Hypothesen und wird eine Aussage treffen für welche Anzahl der 53 Studien diese Hypothesen jeweils zutreffend sind.

1. *Hypothese*
   Die durch die Einleitung der klinischen Studie erwartete Intervention ist identisch mit der tatsächlich durchgeführten Intervention der Methodik.

2. *Hypothese*
   Das, durch die definierten Endpunkte in der Methodik, erwartete Ergebnis der klinischen Studie stimmt mit dem beobachteten Ergebnis in der Diskussion überein.

3. *Hypothese*
   Das beobachtete Ergebnis in der Diskussion einer klinischen Studie entspricht dem beobachteten Ergebnis im Abstract.

4. *Hypothese*
   Die durch den Titel erwartete Intervention der klinischen Studie entspricht der tatsächlich durchgeführten Intervention in der Methodik dieser Studie.

5. *Hypothese*
   Die formulierte Fragestellung im Titel einer klinischen Studie wird in der Diskussion dieser Studie durch ein beobachtetes Ergebnis beantwortet.

Jeder Hypothese ist ein Schlüsselpunkt in einer der ersten drei Ebenen zugeordnet. Durch die Anzahl der inhaltlichen Übereinstimmungen des jeweiligen Schlüsselpunktes kann eine Aussage getroffen werden, für welche Anzahl der 53 Studien die jeweilige Hypothese zutrifft. Die folgende Tabelle zeigt die Zuordnung der Hypothesen zu dem entsprechenden Schlüsselpunkt.

Tabelle 6: Die fünf Hypothesen

Die Hypothese...	*…entspricht der Anzahl der inhaltlichen Übereinstimmungen im Schlüsselpunkt…*
H1	…*DIE INTERVENTION* in der ersten Ebene
H2	…*DAS ERGEBNIS* in der zweiten Ebene E2A
H3	…*DAS ERGEBNIS* in der zweiten Ebene E2B
H4	…*DIE INTERVENTION* in der dritten Ebene E3A
H5	…*DAS ERGEBNIS* in der dritten Ebene E3B

Der zweite Auswertungsschritt: Die Identifikation der Problemfelder

Der zweite Auswertungsschritt überprüft die inhaltliche Konsistenz der fünf Textstellen Titel, Einleitung, Methodik, Diskussion und Abstract innerhalb einer Studie. Die erste Ebene der ICCT vergleicht den Inhalt der Einleitung mit dem Inhalt der Methodik. Die zweite Ebene der ICCT führt denselben Vergleich zwischen der Methodik, der Diskussion und dem Abstract. Die dritte Ebene der ICCT führt diesen Vergleich zwischen Titel, Methodik und Diskussion.

Der zweite Auswertungsschritt zählt also die Anzahl der Studien, die in der zu untersuchenden Ebene (der ersten, zweiten oder dritten Ebene) in mindestens einem Schlüsselpunkt dieser Ebene inhaltlich nicht übereinstimmen. Dadurch kann festgestellt werden, welche Textstellen am häufigsten inhaltlich nicht übereinstimmen.

Tabelle 7: Identifikation der Problemfelder

***Die Ebene**... (...vergleicht die inhaltliche Übereinstimmung der Textstellen...)*	*Die Anzahl der Studien, die an mindestens einem Schlüsselpunkt dieser Ebene nicht übereingestimmt haben.*
E1 (*... Einleitung und Methodik)*	x von 53 Studien
E2 *(...Methodik, Diskussion und Abstract)*	x von 53 Studien
E3 *(...Titel, Methodik und Diskussion)*	x von 53 Studien

Der dritte Auswertungsschritt: Die Analyse der Handlungsfelder

Der dritte Auswertungsschritt prüft die inhaltliche Konsistenz jedes einzelnen Schlüsselpunktes der ersten drei Ebenen. Er ermöglicht eine Aussage darüber, welcher Schlüsselpunkt inhaltlich am wenigsten häufig übereingestimmt hat. Dadurch kann in der Bewertung der Ergebnisse der ICCT gezielt aufgezeigt werden, an welchen Schlüsselpunkten in einer klinischen Studie der meiste Handlungsbedarf besteht.
Dazu wird die Anzahl derjenigen Studien gezählt die im entsprechenden Schlüsselpunkt übereinstimmen. Ebenso wird die Anzahl der Studien gezählt, die im entsprechenden Schlüsselpunkt nicht übereinstimmen. Auch wird die Anzahl der Studien angegeben, die am entsprechenden Schlüsselpunkt möglicherweise übereinstimmen.

Tabelle 8: Analyse der Handlungsfelder

Die Schlüsselpunkte in der ersten Ebene (Einleitung vs. Methodik)		*Anzahl der inhaltlichen Übereinstimmungen*	*Anzahl der möglichen inhaltlichen Übereinstimmungen*	*Anzahl der fehlenden inhaltlichen Übereinstimmungen*
1.	**Die Studiengruppe**	x von n=53 Studien	x von n=53 Studien	x von n=53 Studien
2.	**Die Intervention**	x von n=53 Studien	x von n=53 Studien	x von n=53 Studien
3.	**Das Ziel der Intervention**	x von n=53 Studien	x von n=53 Studien	x von n=53 Studien
Die Schlüsselpunkte in der zweiten Ebene E2A (Methodik vs. Diskussion)		*Anzahl der inhaltlichen Übereinstimmungen*	*Anzahl der möglichen inhaltlichen Übereinstimmungen*	*Anzahl der fehlenden inhaltlichen Übereinstimmungen*
1.	**Die Studiengruppe**	x von n=53 Studien	x von n=53 Studien	x von n=53 Studien
2.	**Das Ergebnis**	x von n=53 Studien	x von n=53 Studien	x von n=53 Studien
Die Schlüsselpunkte in der zweiten Ebene E2B (Diskussion vs. Abstract)		*Anzahl der inhaltlichen Übereinstimmungen*	*Anzahl der möglichen inhaltlichen Übereinstimmungen*	*Anzahl der fehlenden inhaltlichen Übereinstimmungen*
1.	**Die Studiengruppe**	x von n=53 Studien	x von n=53 Studien	x von n=53 Studien
2.	**Das Ergebnis**	x von n=53 Studien	x von n=53 Studien	x von n=53 Studien
Der Schlüsselpunkt in der dritten Ebene E3A (Titel vs. Methodik)		*Anzahl der inhaltlichen Übereinstimmungen*	*Anzahl der möglichen inhaltlichen Übereinstimmungen*	*Anzahl der fehlenden inhaltlichen Übereinstimmungen*

1.	**Die Intervention**	x von n=53 Studien	x von n=53 Studien	x von n=53 Studien
Der Schlüsselpunkte in der dritten Ebene E3B (Titel vs. Diskussion)		*Anzahl der inhaltlichen Übereinstimmungen*	*Anzahl der möglichen inhaltlichen Übereinstimmungen*	*Anzahl der fehlenden inhaltlichen Übereinstimmungen*
1.	**Das Ergebnis**	x von n=53 Studien	x von n=53 Studien	x von n=53 Studien

Der vierte Auswertungsschritt: Die speziellen Analysen

Mit Hilfe der vorangegangenen Auswertungsschritte ist es möglich Problemfelder zu identifizieren. Das bedeutet, es kann eine Aussage getroffen werden, welche der ersten drei Ebenen inhaltlich am häufigsten nicht übereingestimmt hat. Durch die Identifikation der Handlungsfelder kann dieselbe Aussage für die einzelnen Schlüsselpunkte getroffen werden. Allerdings kann noch keine Aussagen darüber getroffen werden, warum die einzelnen Schlüsselpunkte in ihrem Inhalt nicht übereingestimmt haben. Aus diesem Grund werden im vierten und letzten Auswertungsschritt nochmals spezielle Analysen durchgeführt, die unter anderem diese Frage beantworten können. Auch die Auswertung der vierten Ebene erfolgt in diesem Auswertungsschritt.

Die Ursachen einer fehlenden Übereinstimmung in den Schlüsselpunkten DAS ZIEL DER INTERVENTION und DAS ERGEBNIS (Die Analyse der Endpunkte)

Diese Analyse untersucht den Schlüsselpunkt *DAS ZIEL DER INTERVENTION* aus der ersten Ebene und den Schlüsselpunkt *DAS ERGEBNIS* aus der zweiten Ebene genauer. Es werden die Ursachen aufgezeigt, warum die Schlüsselpunkte in diesen Ebenen nicht übereingestimmt haben. Für den Schlüsselpunkt *DAS ERGEBNIS* der zweiten Ebene erfolgte eine getrennte Aufstellung für die Ebene E2A und die Ebene E2B.

In der ersten Ebene, dem Vergleich von Einleitung und Methodik, bestehen für den Schlüsselpunkt *DAS ZIEL DER INTERVENTION* drei mögliche Ursachen für eine inhaltliche Nichtübereinstimmung.

1. *Für eine formulierte Fragestellung in der Einleitung der Studie findet sich kein entsprechender definierter Endpunkt in der Methodik dieser Studie.*

2. *Für einen definierten Endpunkt in der Methodik der Studie findet sich in der Einleitung keine entsprechend formulierte Fragestellung.*

3. *Für eine formulierte Fragestellung in der Einleitung der Studie findet sich kein entsprechender definierter Endpunkt in der Methodik der Studie und gleichzeitig findet sich für einen definierten Endpunkt in der Methodik der Studie keine entsprechend formulierte Fragestellung in der Einleitung.*

In der zweiten Ebene E2A, dem Vergleich von Methodik und Diskussion, bestehen für den Schlüsselpunkt *DAS ERGEBNIS* fünf mögliche Ursachen für eine inhaltliche Nichtübereinstimmung.

1. *Für den primären Endpunkt aus der Methodik der Studie findet sich kein beobachtetes Ergebnis in der Diskussion.*

2. *Für einen sekundären Endpunkt aus der Methodik der Studie findet sich kein beobachtetes Ergebnis in der Diskussion.*

3. *Für einen zusätzlichen Endpunkt aus der Methodik der Studie findet sich kein beobachtetes Ergebnis in der Diskussion.*

4. *Für ein beobachtetes Ergebnis aus der Diskussion der Studie findet sich kein definierter Endpunkt in der Methodik.*

5. *Eine Zuordnung zu den bisher genannten Ursachen ist nicht möglich*

In der zweiten Ebene E2B, dem Vergleich von Diskussion und Abstract, bestehen für den Schlüsselpunkt *DAS ERGEBNIS* ebenfalls fünf mögliche Ursachen für eine inhaltliche Nichtübereinstimmung.

1. *Das für den primären Endpunkt in der Diskussion der Studie beobachtete Ergebnis fehlt im Abstract.*

2. *Das für einen sekundären Endpunkt in der Diskussion der Studie beobachtete Ergebnis fehlt im Abstract.*

3. *Das für einen zusätzlichen Endpunkt in der Diskussion der Studie beobachtete Ergebnis fehlt im Abstract.*

4. *Für ein Ergebnis im Abstract der Studie findet sich kein beobachtetes Ergebnis in der Diskussion.*

5. *Eine Zuordnung zu den bisher genannten Ursachen ist nicht möglich*

Es wird gezählt für welche Anzahl an Studien eine der oben genannten Ursachen der Grund für eine inhaltliche Nichtübereinstimmung des jeweiligen Schlüsselpunktes war. Für jede Ursache wird die absolute Anzahl an Studien gezählt, für die diese Ursache zutrifft. Diese Zahl wird dann als Prozentsatz aller 53 Studien angegeben. Zuletzt wird der prozentuale Anteil jeder einzelnen Ursache im Vergleich zu allen Ursachen einer Nichtübereinstimmung an diesem Schlüsselpunkt angegeben.

Tabelle 9: Ursachen für eine fehlende Übereinstimmung im Schlüsselpunkt ZIEL DER INTERVENTION in der ersten Ebene

Legende für Tabelle 9:
[M*]: Für eine formulierte Fragestellung in der Einleitung findet sich kein definierter Endpunkt in der Methodik.
[E*]: Für einen definierten Endpunkt aus der Methodik der findet sich keine entsprechende Fragestellung in der Einleitung
[M*] und [E*]: Beide Ursachen treffen zu

Die drei Ursachen für eine Nichtübereinstimmung des Schlüsselpunktes „Das Ziel der Intervention“ in der ersten Ebene (Einleitung vs. Methodik)	*Die Anzahl der Studien, in der diese Ursache zu finden ist*	*Der prozentuale Anteil dieser Studien an der Gesamtzahl von 53 Studien*
[M]*	x	x von 53
[E]*	x	x von 53
[M] und [E*]*	x	x von 53

Tabelle 10: Ursachen für eine fehlende Übereinstimmung im Schlüsselpunkt DAS ERGEBNIS der zweiten Ebene E2A

Legende für Tabelle 10:

[M]:** Für ein beobachtetes Ergebnis aus der Diskussion findet sich kein definierter Endpunkt in der Methodik

P : Für den primären Endpunkt aus der Methodik findet sich kein beobachtetes Ergebnis in der Diskussion

S : Für einen sekundären Endpunkt aus der Methodik findet sich kein beobachtetes Ergebnis in der Diskussion

A : Für einen zusätzlichen Endpunkt aus der Methodik findet sich kein beobachtetes Ergebnis in der Diskussion

*** :** Die Zuordnung zu den bisher genannten Ursachen ist nicht möglich

Die fünf Ursachen für eine Nichtübereinstimmung des Schlüsselpunktes „Das Ergebnis“ in der zweiten Ebene E2A (Methodik vs. Diskussion)	*Die Anzahl der Studien, in der diese Ursache zu finden ist*	*Der prozentuale Anteil dieser Studien an der Gesamtzahl von 53 Studien*
*[M**]*	x	x von 53
P	x	x von 53
S	x	x von 53
A	x	x von 53
*Nicht zuzuordnen **	x	x von 53

Tabelle 11: Ursachen für eine fehlende Übereinstimmung im Schlüsselpunkt DAS ERGEBNIS der zweiten Ebene E2B

Legende für Tabelle 11:

[D] :** Für ein beobachtetes Ergebnis im Abstract findet sich kein beobachtetes Ergebnis in der Diskussion

P : Das für den primären Endpunkt beobachtete Ergebnis in der Diskussion fehlt im Abstract

S : Das für einen sekundären Endpunkt beobachtete Ergebnis in der Diskussion fehlt im Abstract

A : Das für einen zusätzlichen Endpunkt beobachtete Ergebnis in der Diskussion fehlt im Abstract

**** :** Die Zuordnung zu den bisher genannten Ursachen ist nicht möglich

Die fünf Ursachen für eine Nichtübereinstimmung des Schlüsselpunktes „Das Ergebnis“ in der zweiten Ebene E2B (Diskussion vs. Abstract)	*Die Anzahl der Studien, in der diese Ursache zu finden ist*	*Der prozentuale Anteil dieser Studien an der Gesamtzahl von 53 Studien*
*[D**]*	x	x von 53
P	x	x von 53
S	x	x von 53
A	x	x von 53
*Nicht zuzuordnen ***	x	x von 53

Andere Ursachen für eine fehlende Übereinstimmung im Schlüsselpunkte DAS ERGEBNIS

Für den Schlüsselpunkt DAS ERGEBNIS der zweiten Ebene kann die Ursache für eine Nichtübereinstimmung dieses Schlüsselpunktes nicht immer einer der ersten vier Ursachen, wie oben genannt, zugeordnet werden. Aus diesem Grund werden in diesem Auswertungsschritt alle anderen Ursachen aufgezählt und beschrieben, die zu einer fehlenden inhaltlichen Konsistenz des Schlüsselpunktes DAS ERGEBNIS in der zweiten Ebene geführt haben. Eine getrennte Aufstellung in die Ebenen E2A und E2B erfolgt nicht.

Die Auswertung der vierten Ebene der ICCT - Der primäre Endpunkt im Abstract

Die vierte Ebene der ICCT vergleicht keine Textstellen wie die vorangegangenen drei Ebenen. Aus diesem Grund fehlt die vierte Ebene in der Auswertungspyramide. Die vierte Ebene vergleicht die Schlussfolgerung im Abstract mit dem primären Endpunkt einer Studie. Dabei stellt sie die Frage, ob sich die Schlussfolgerung im Abstract auf das beobachtete Ergebnis für den primären Endpunkt bezieht.

Es wird die Anzahl der Studien gezählt, in denen sich die Schlussfolgerung im Abstract direkt auf den primären Endpunkt bezieht. Auch die Anzahl der Studien, in denen sich die Schlussfolgerung im Abstract nicht auf den primären Endpunkt bezieht, wird gezählt. Da nicht immer zweifelsfrei entschieden werden kann, ob sich die Schlussfolgerung im Abstract direkt auf den primären Endpunkt bezieht, wird auch die Anzahl der Studien genannt, in denen sich die Schlussfolgerung im Abstract möglicherweise auf den primären Endpunkt bezieht.

Tabelle 12: Bezugnahme der Conclusion im Abstract auf den primären Endpunkt

Anzahl der Studien mit direktem Bezug der Schlussfolgerung auf den primären Endpunkt	*Prozentualer Anteil dieser Studien an allen 53 Studien*	*Anzahl der Studien mit möglicherweise vorhandenem Bezug der Schlussfolgerung auf den primären Endpunkt*	*Prozentualer Anteil dieser Studien an allen 53 Studien*	*Anzahl der Studien mit fehlendem Bezug der Schlussfolgerung auf den primären Endpunkt*	*Prozentualer Anteil dieser Studien an allen 53 Studien*
x Studien	y Prozent von 53	x Studien	y Prozent von 53	x Studien	y Prozent von 53

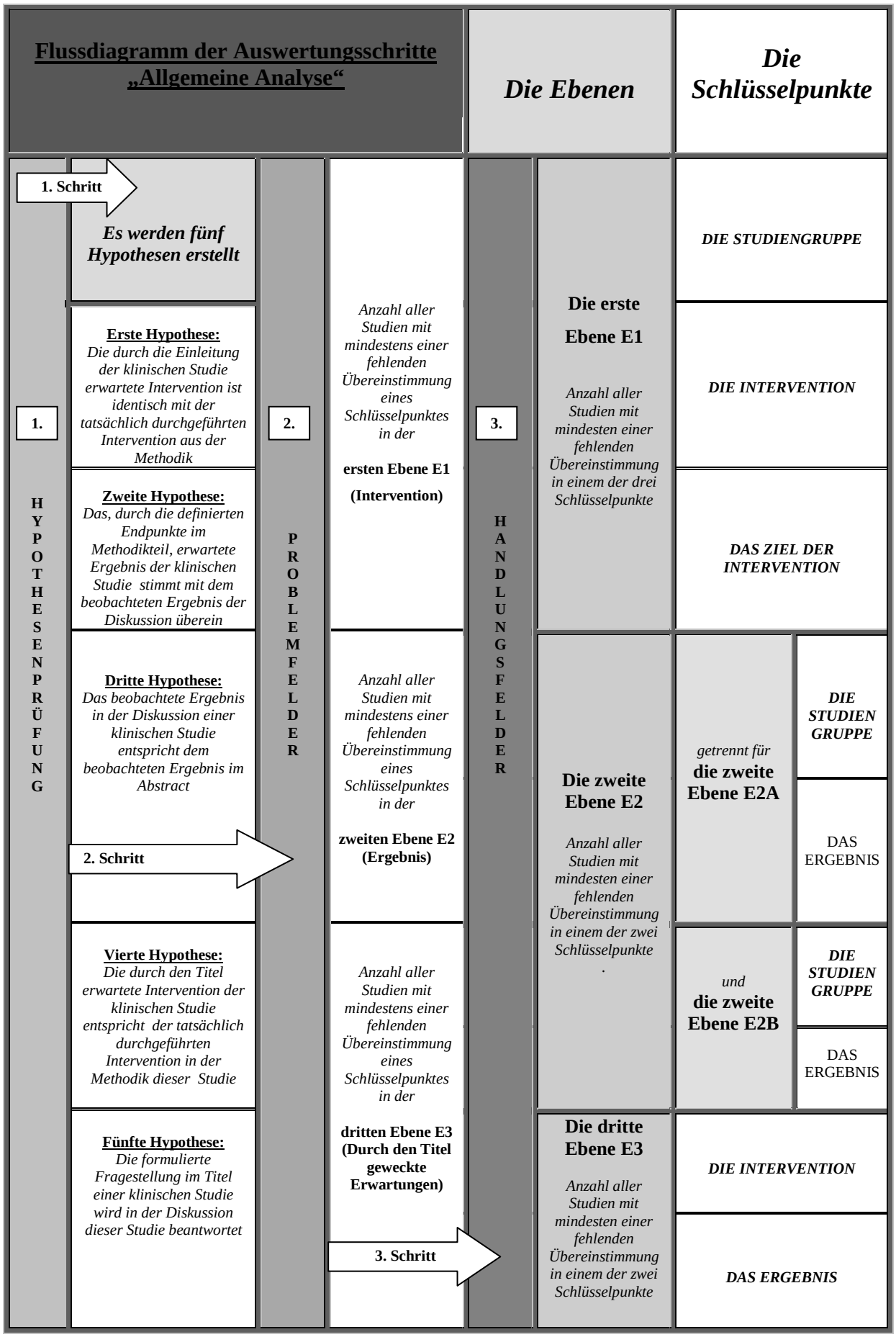

Abbildung 3: Flussdiagramm zu den Auswertungsschritten der „Allgemeinen Analyse“

Im vierten Schritt werden die Ursachen geprüft, warum der Schlüsselpunkt DAS ERGEBNIS in der zweiten Ebene E2 inhaltlich nicht übereingestimmt hat.

4. Schritt

4. SPEZIELLE ANALYSEN

Die zweite Ebene E2A – Der Schlüsselpunkt *DAS ERGEBNIS* Alle Studien mit einer fehlenden Übereinstimmung im Schlüsselpunkt *DAS ERGEBNIS in der zweiten Ebene E2A* werden in diesem Schritt ausgewertet	**Die zweite Ebene E2B – Der Schlüsselpunkt *DAS ERGEBNIS*** Alle Studien mit einer fehlenden Übereinstimmung im Schlüsselpunkt *DAS ERGEBNIS in der zweiten Ebene E2B* werden in diesem Schritt ausgewertet
Der Schlüsselpunkt stimmt am ***primären Endpunkt*** nicht überein	Der Schlüsselpunkt stimmt nicht überein da das beobachtete Ergebnis für ***den primären Endpunkt*** aus der Diskussion im Abstract fehlt
Der Schlüsselpunkt stimmt an einem ***sekundären Endpunkt*** nicht überein	Der Schlüsselpunkt stimmt nicht überein da das beobachtete Ergebnis für ***einen sekundären Endpunkt*** aus der Diskussion im Abstract fehlt
Der Schlüsselpunkt stimmt an einem ***zusätzlichen Endpunkt*** nicht überein	Der Schlüsselpunkt stimmt nicht überein da das beobachtete Ergebnis für ***einen zusätzlichen Endpunkt*** aus der Diskussion im Abstract fehlt
Der Schlüsselpunkt stimmt nicht überein **da einem beobachteten Ergebnis aus der Diskussion kein entsprechender Endpunkt in der Methodik** zugeordnet werden kann	Der Schlüsselpunkt stimmt nicht überein **da für ein Ergebnis aus dem Abstract das entsprechende beobachtete Ergebnis in der Diskussion** fehlt
Eine Zuordnung zu einer der vier oben genannten Ursachen ist nicht möglich	Eine Zuordnung zu einer der vier oben genannten Ursachen ist nicht möglich

Abbildung 4: Flussdiagramm zum Auswertungsschritt „Spezielle Analyse der Endpunkte“

3. Ergebnisse

Die ICCT untersuchte insgesamt 53 Studien aus drei Journals auf ihre inhaltliche Konsistenz Die Studien sind im Literaturverzeichnis angegeben. [79] bis [131] Dabei wurden aus dem *New England Journal of Medicine* 27 Studien untersucht. 17 der untersuchten Studien stammten aus *The Lancet*. Die restlichen neun Studien stammten aus dem *American Journal of Obstetrics and Gynecology*. Die ungleiche Verteilung der Studien auf die drei Journals erklärt sich aus den unterschiedlichen Zugriffsmöglichkeiten auf die jeweilige Volltextversion der Studien in den Internetauftritten der Journals. Während der Volltextzugriff im *New England Journal of Medicine* auf alle publizierten Studien möglich war, beschränkte er sich beim *American Journal of Obstetrics and Gynecology* auf die Studien der letzten 12 Monate.

Die durchschnittliche Bearbeitungszeit einer Studie betrug circa 1,5 Stunden. Dabei wurde die Studie im ersten Schritt komplett durchgelesen. Im zweiten Schritt wurden die Schlüsselpunkte in den Ebenen identifiziert und auf den Auswertungsbogen übertragen. Im letzten Schritt wurde anhand der definierten Kriterien die inhaltliche Konsistenz der Schlüsselpunkte geprüft.

Nachdem alle 53 Studien auf ihre inhaltliche Konsistenz überprüft worden waren, erfolgte die Auswertung der Ergebnisse anhand des zuvor beschriebenen Auswertungsschemas. Diese Ergebnisse werden nun im Einzelnen vorgestellt.

3.1 Das Ergebnis für die fünf Hypothesen

1. Die durch die Einleitung der klinischen Studie erwartete Intervention ist identisch mit der tatsächlich durchgeführten Intervention der Methodik.

2. Das, durch die definierten Endpunkte im Methodikteil, erwartete Ergebnis der klinischen Studie stimmt mit dem beobachteten Ergebnis in der Diskussion überein.

3. Das beobachtete Ergebnis in der Diskussion einer klinischen Studie entspricht dem beobachteten Ergebnis im Abstract.

4. Die durch den Titel erwartete Intervention der klinischen Studie entspricht der tatsächlich durchgeführten Intervention in der Methodik dieser Studie.

5. Die formulierte Fragestellung im Titel einer klinischen Studie wird in der Diskussion dieser Studie beantwortet.

Die erste Hypothese sagt aus, dass die Einleitung einer klinischen Studie die Intervention in der Studiengruppe genauso beschreibt, wie sie dann in der Methodik durchgeführt wird. Diese Hypothese traf für 32 der 53 Studien zu. Das entspricht 60,4 Prozent aller Studien

Die zweite Hypothese geht von der Annahme aus, dass für jeden definierten Endpunkt in der Methodik einer klinischen Studie ein entsprechendes Ergebnis in der Diskussion beobachtet wird. Diese Annahme traf nur für 11 der untersuchten 53 Studien zu. Das entspricht 20,8 Prozent der untersuchten Studien.

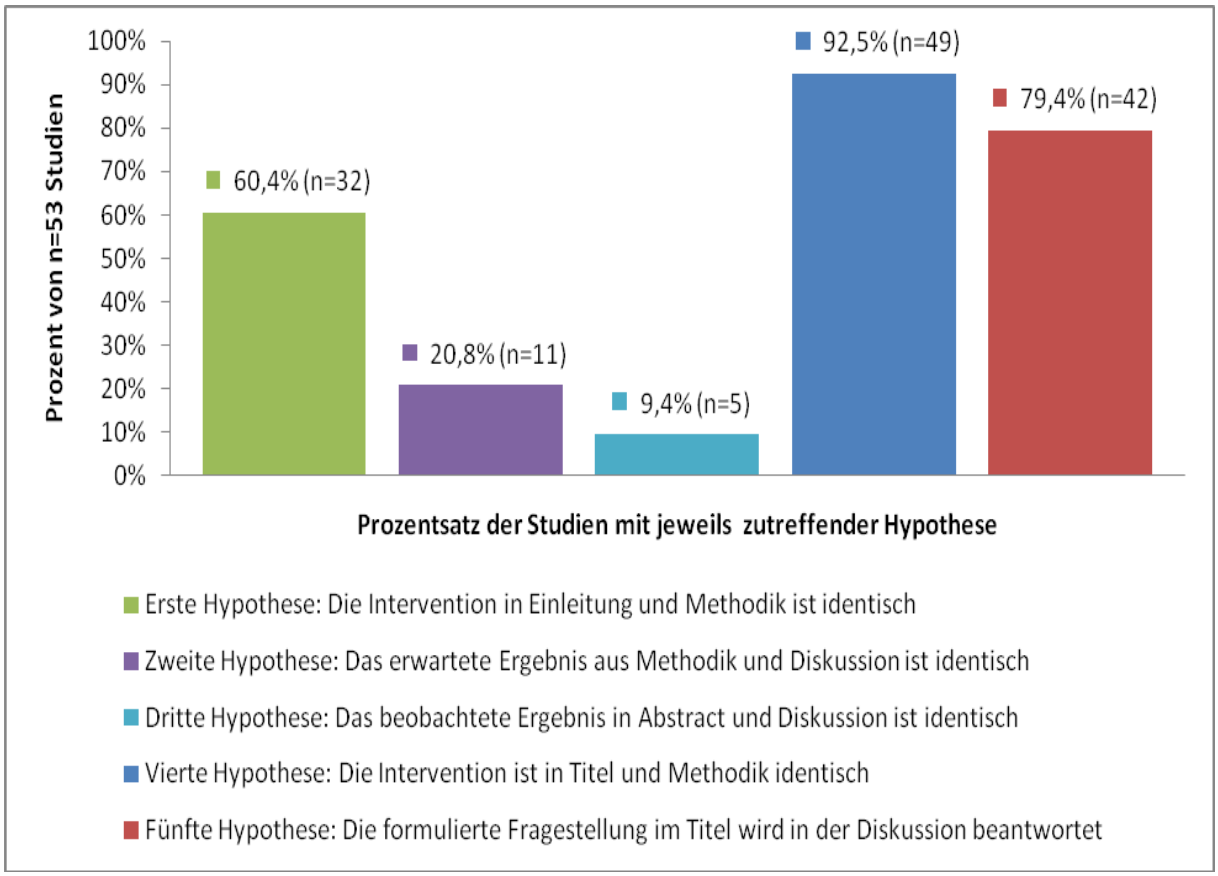

Abbildung 5: Ergebnis für die Prüfung der fünf Hypothesen

Die dritte Hypothese sagt aus, dass jedes beobachtete Ergebnis aus der Diskussion einer klinischen Studie in Kurzform auch im Abstract aufgeführt wird. Nach Auswertung der 53 Studien traf diese Annahme nur für fünf dieser 53 Studien zu. Das entspricht 9,4 Prozent aller untersuchter Studien.

In der vierten Hypothese geht die ICCT davon aus, dass die Intervention im Titel einer klinischen Studie der tatsächlichen Intervention in der Methodik dieser Studie entspricht. Diese Hypothese traf für 49 der 53 Studien zu. Das entspricht 93,5 Prozent aller Studien.

Als letzte Hypothese geht die ICCT vom Ansatz aus, dass die formulierte Fragestellung im Titel einer klinischen Studie in der Diskussion durch ein beobachtetes Ergebnis beantwortet wird. Dies traf für 42 der 53 Studien zu. Das entspricht 79,4 Prozent.

Die Prozentzahlen der vierten und fünften Hypothese berechnen sich auf der Grundlage der 53 untersuchten Studien. Nicht jede Studie formuliert die Intervention bereits im Titel oder führt im Titel eine Fragestellung an die Studie ein. Somit müssen die oben genannten Prozentzahlen für die vierte und fünfte Hypothese entsprechend korrigiert werden.
Vier Studien beschrieben keine Intervention im Titel. Zieht man diese vier Studien von der Gesamtzahl der 53 Studien ab, so trifft die vierte Hypothese für 100 Prozent aller Studien zu, die eine Intervention im Titel formuliert haben.
Sieben Studien hatten keine Fragestellung im Titel formuliert. Zieht man wiederum diese sieben Studien von der Gesamtzahl der 53 Studien ab, so trifft die fünfte Hypothese für 89,4 Prozent derjenigen Studien zu, die im Titel eine Fragestellung formuliert hatten.

3.2 Das Ergebnis zur Identifikation der Problemfelder

Bei der Indentifikation der Problemfelder überprüft die ICCT die ersten drei Ebenen. Es wird untersucht, wie oft die 53 Studien in mindestens einem Schlüsselpunk der jeweiligen Ebene inhaltlich nicht übereingestimmt haben.

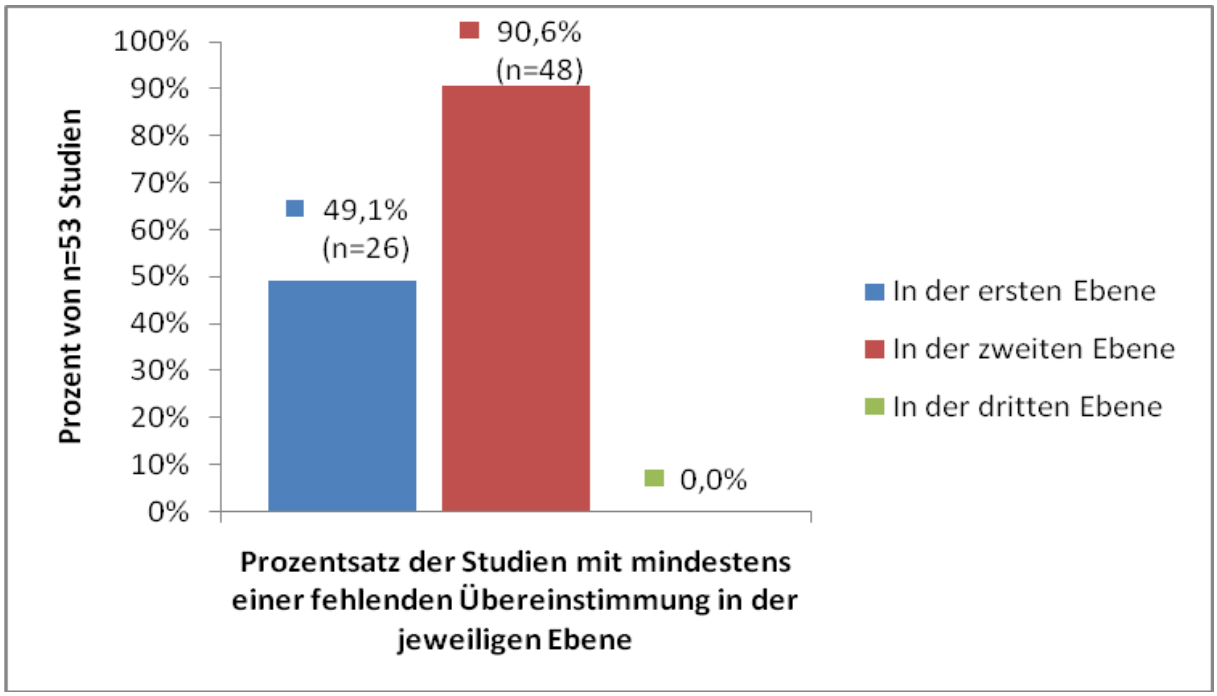

Abbildung 6: Die Identifikation der Problemfelder

In der **ersten Ebene,** dem Vergleich zwischen der Einleitung und der Methodik einer klinischen Studie stimmten 26 der 53 Studien an mindestens einem der drei Schlüsselpunkte dieser Ebene nicht überein. Das bedeutet, dass sich die Einleitung bei 49,1 Prozent aller geprüfter Studien in bestimmten Punkten von der Methodik derselben Studie unterschied.

Die **zweite Ebene** vergleicht die Methodik mit der Diskussion und die Diskussion mit dem Abstract. Hier stimmten 48 der untersuchten 53 Studien an mindestens einem der Schlüsselpunkt dieser Ebene nicht überein. Somit hatten in 90,6 Prozent aller untersuchter Studien die Textstellen Methodik und Diskussion oder Diskussion und Abstract in Teilen einen unterschiedlichen Inhalt.

In der **dritten Ebene** vergleicht die ICCT den Titel mit der Methodik und der Diskussion. Bei der Auswertung der Ergebnisse fand sich keine Studie, die in dieser Ebene inhaltlich nicht übereingestimmt hat. Das bedeutet, dass in keiner Studie, die für die ICCT gelesen wurde, die formulierte Fragestellung aus dem Titel in der Diskussion nicht beantwortet wurde. Ebenso unterschied sich die beschriebene Intervention im Titel in keiner Studie von der tatsächlich durchgeführten Intervention aus der Methodik.

3.3 Das Ergebnis für die Analyse der Handlungsfelder

Die Analyse der Handlungsfelder zählt die Anzahl der Studien mit einer fehlenden inhaltlichen Übereinstimmung getrennt nach den einzelnen Schlüsselpunkten der ersten drei Ebenen.

Für die **erste Ebene** wurden drei Schlüsselpunkte definiert: *DIE STUDIENGRUPPE, DIE INTERVENTION* und *DAS ZIEL DER INTERVENTION*.

Verglich man die vorgestellte **Studiengruppe** aus der Einleitung mit der Studiengruppe in der Methodik so waren diese in 40 der untersuchten 53 Studien identisch. Definitiv nicht identisch waren diese beiden Gruppen in drei der untersuchten 53 Studien. [84][93][111] In zehn Studien war eine Entscheidung nicht möglich.

Die geplante **Intervention** aus der Einleitung entsprach bei 32 der 53 Studien exakt der durchgeführten Intervention in der Methodik. In sechs der 53 Studien beschrieben diese beiden Textstellen jedoch eine unterschiedliche Interventionen. [80][87][89][109][111][122]

In 15 Studien konnte keine Entscheidung getroffen werden. Das lag hauptsächlich daran, dass eine Intervention gegen einen Placeboarm geprüft wurde. Allerdings wurde das in der Einleitung so nicht explizit erwähnt. Nach den im Anhang aufgeführten Definitionen galt dieser Fall nicht als fehlende inhaltliche Übereinstimmung, sondern wurde als mögliche inhaltliche Übereinstimmung gewertet.

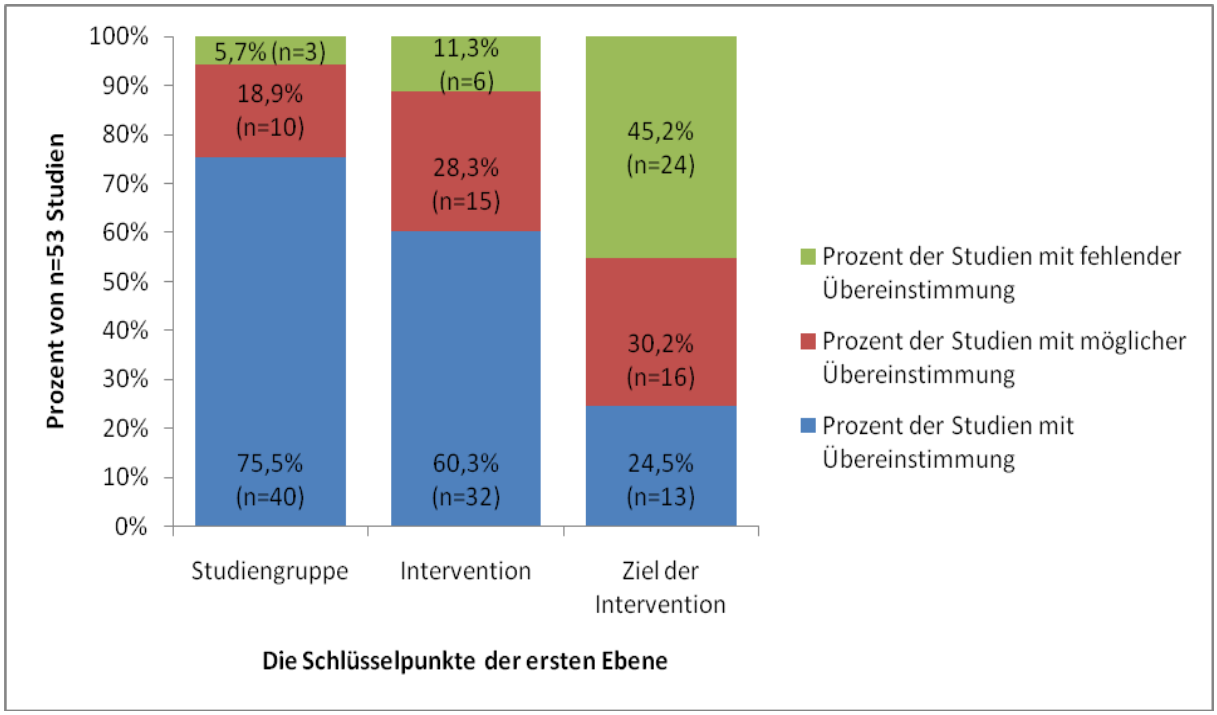

Abbildung 7: Analyse der Handlungsfelder - Das Ergebnis für die Schlüsselpunkte der ersten Ebene -

Der Vergleich der **Fragestellungen (Ziel der Intervention)** an die Studie aus der Einleitung mit den tatsächlich definierten Endpunkten in der Methodik ergab, dass nur 13 der untersuchten 53 Studien hier komplett übereinstimmten. In 24 Studien stimmten die Fragestellungen aus der Einleitung nicht in allen Teilen mit den definierten Endpunkten in der Methodik überein. Bei 16 Studien war eine Entscheidung nicht möglich.

Für die **zweite Ebene** definierte die ICCT zwei Schlüsselpunkte: *DIE STUDIENGRUPPE* und *DAS ERGEBNIS*. Da in der zweiten Ebene insgesamt drei Textstellen verglichen werden, die Methodik mit der Diskussion und die Diskussion mit dem Abstract, wurde die zweite Ebene in die Ebenen E2A und E2B getrennt.

In der **Ebene E2A** wurden die Textstellen Methodik und Diskussion miteinander verglichen Verglich man die **Studiengruppe** in der Methodik mit der Gruppe aus der Diskussion, für die anhand der beobachteten Ergebnisse eine Aussage getroffen wurde, so waren diese beiden Gruppen in 33 der untersuchten 53 Studien identisch. In 15 Studien galten die gefundenen

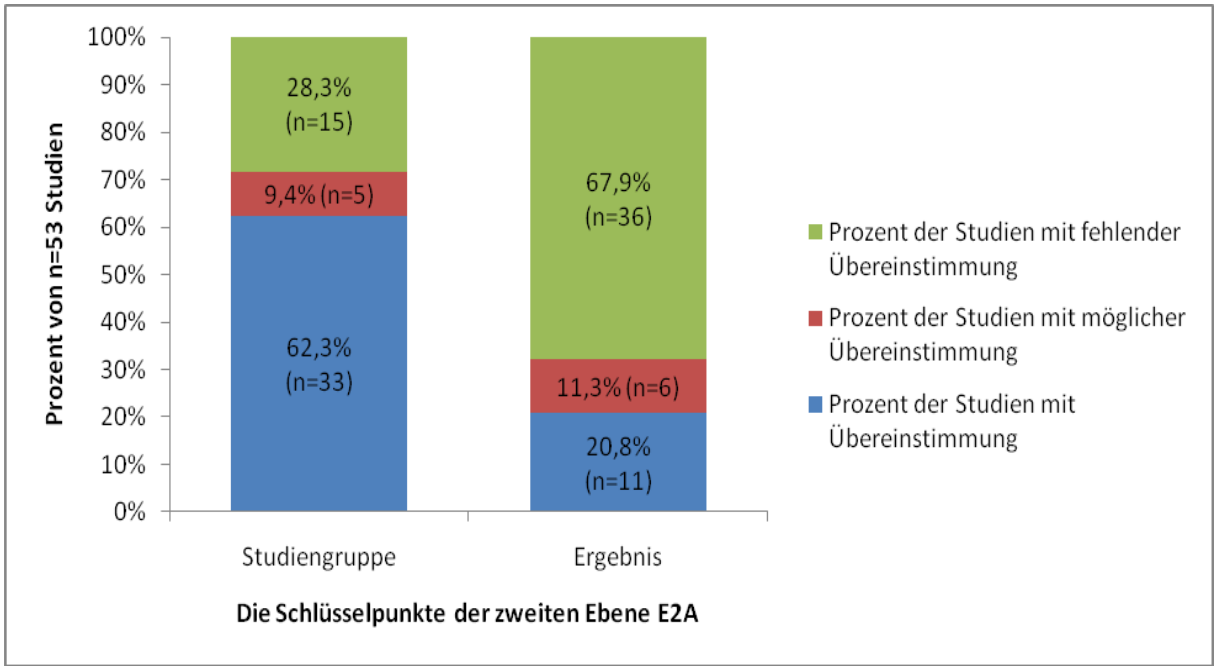

Abbildung 8: Analyse der Handlungsfelder - Das Ergebnis für die Schlüsselpunkte der zweiten Ebene E2A –

Ergebnisse in der Diskussion für eine Gruppe, die nicht exakt der ursprünglichen Studiengruppe in der Methodik entsprach.[80][82][84][85][90][92][95][96][104][107][111][118][122][123][124] Für fünf Studien konnte keine Entscheidung getroffen werden.

Der Vergleich der definierten **Endpunkte (Ergebnis)** aus der Methodik mit den dafür beobachteten Ergebnissen in der Diskussion ergab, dass diese nur in 11 der 53 Studien inhaltlich komplett übereinstimmten. In 36 Studien entsprachen die definierten Endpunkte der Methodik nicht in allen Punkten den beobachteten Ergebnissen in der Diskussion. Für fünf Studien war keine Entscheidung möglich.

Die **Ebene E2B** verglich die Textstellen Diskussion und Abstract miteinander.

Der Vergleich des Schlüsselpunktes **"Die Studiengruppe"** zwischen diesen beiden Textstellen ergab, dass in 35 der untersuchten 53 Studien beide Textstellen die gleiche Gruppe beschrieben. In neun Studien bezogen sich die beobachteten Ergebnisse in der Diskussion jedoch auf eine Gruppe, die nicht der Studiengruppe im Abstract entsprach.[81][83][84][92][93][95][96][111][126] In neun Studien ist eine Entscheidung nicht möglich.

Verglich man die beobachteten **Ergebnisse** in der Diskussion der Studie mit den Ergebnissen im Abstract, so stimmten diese nur in fünf der insgesamt geprüften 53 Studien komplett überein. Während in zwei Studien keine Entscheidung möglich war, entsprach das Ergebnis

im Abstract in 46 Studien nicht in allen Teilen dem beobachteten Ergebnis der Diskussion. Das entspricht 86,8 Prozent aller untersuchter Studien.

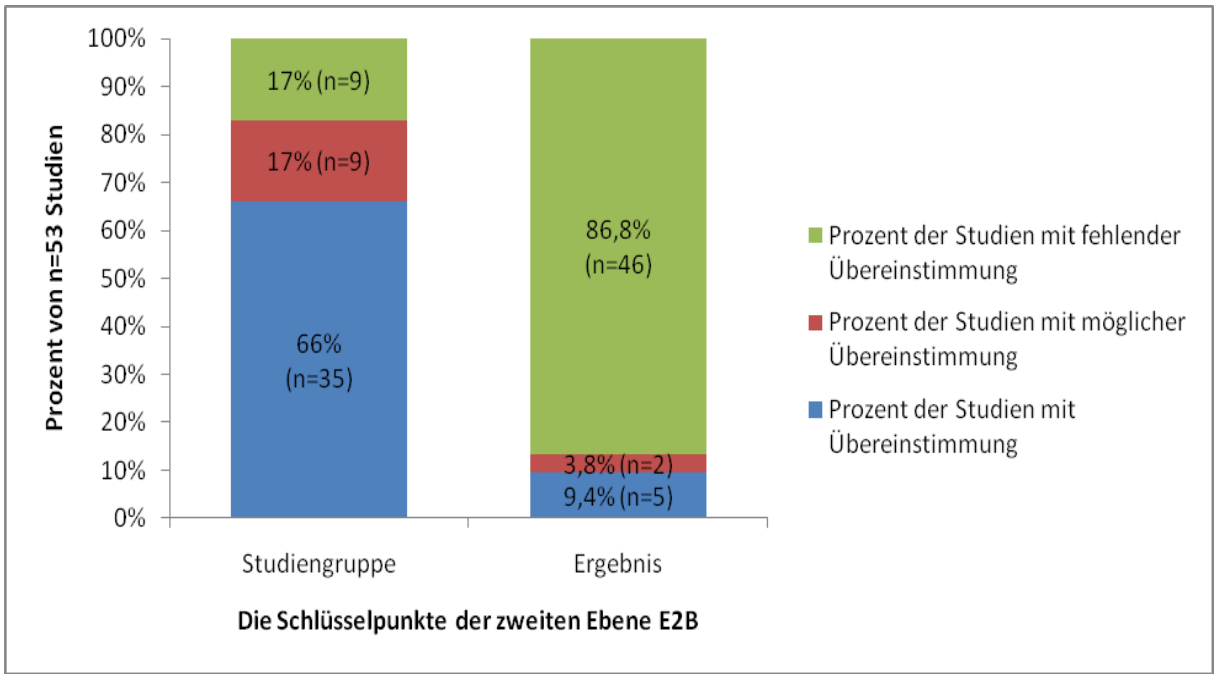

Abbildung 9: Analyse der Handlungsfelder - Das Ergebnis für die Schlüsselpunkte der zweiten Ebene E2B –

In der **dritten Ebene** vergleicht die ICCT den Schlüsselpunkt *DIE INTERVENTION* und den Schlüsselpunkt *DAS ERGEBNIS* zwischen den Textstellen Titel, Methodik und Diskussion. Auch hier erfolgt eine Trennung der dritten Ebene in die Ebenen E3A und E3B.

Die **Ebene E3A** vergleicht den Schlüsselpunkt *DIE INTERVENTION* zwischen dem Titel und der Methodik.

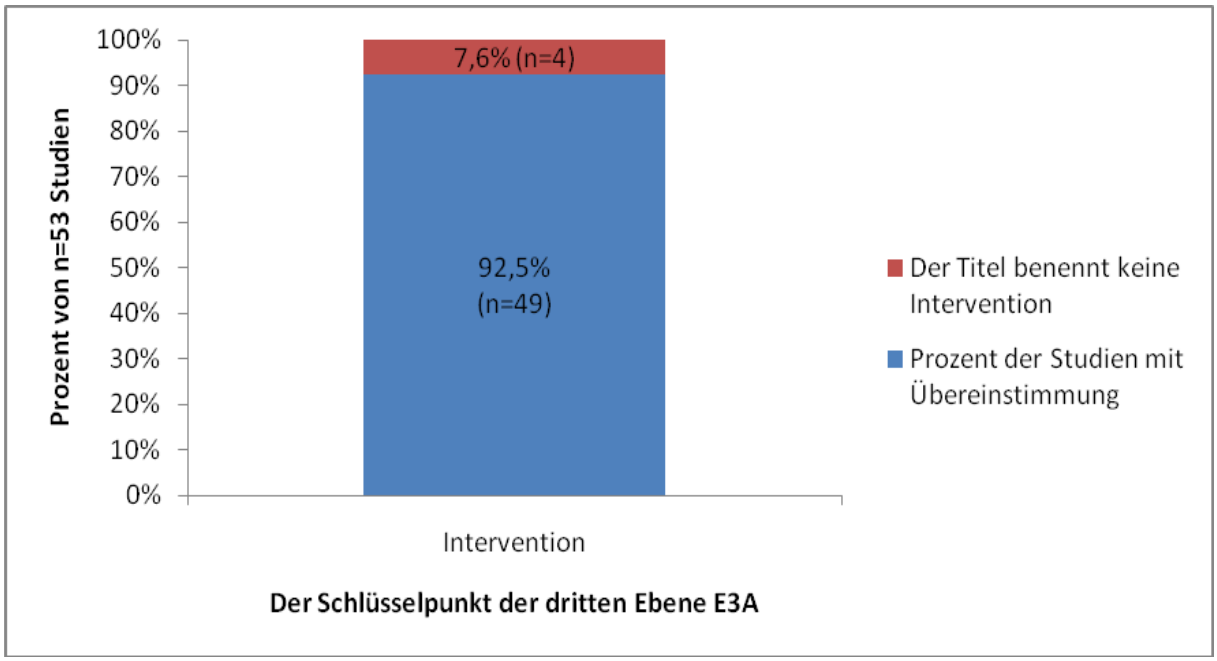

Abbildung 10: Analyse der Handlungsfelder - Das Ergebnis für die Schlüsselpunkte der dritten Ebene E3A –

Benannte der Studientitel eine geplante **Intervention**, dann entsprach diese Intervention in 49 der untersuchten 53 Studien der durchgeführten Intervention in der Methodik. Insgesamt vier der 53 Studien benannten im Titel keine geplante Intervention, sodass alle Studien, die im Titel eine Intervention genannt hatten, diese Intervention auch tatsächlich so durchführten.

Die **Ebene E3B** vergleicht den Schlüsselpunkt *DAS ERGEBNIS* zwischen dem Titel und der Diskussion.

Wurde im Titel einer klinischen Studie eine **Fragestellung (Ergebnis)** an diese Studie formuliert, dann wurde sie in 42 der geprüften 53 Studien durch ein beobachtetes Ergebnis in der Diskussion beantwortet.

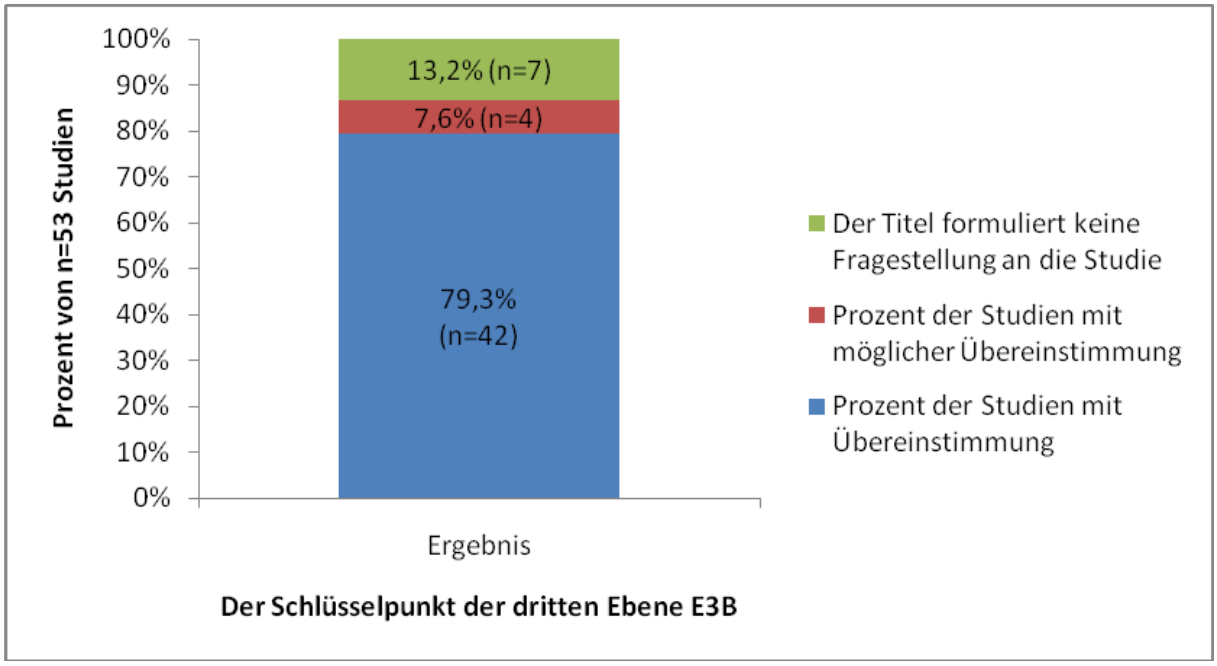

Abbildung 11: Analyse der Handlungsfelder - Das Ergebnis für die Schlüsselpunkte der dritten Ebene E3B –

Sieben der 53 Studien hatten im Titel keine Fragestellung formuliert. In vier Studien war eine Entscheidung nicht möglich. Das bedeutet, dass insgesamt 89,4 Prozent aller geprüfter Studien die Fragestellung im Titel durch ein beobachtetes Ergebnis in der Diskussion beantwortetet hatten.

3.4 Das Ergebnis für die Analyse der Endpunkte

Diese Analyse untersuchte den Schlüsselpunkt *DAS ZIEL DER INTERVENTION* der ersten Ebene und den Schlüsselpunkt *DAS ERGEBNIS* der zweiten Ebene genauer. Es sollten die Ursachen gefunden werden, warum diese Schlüsselpunkte nicht übereingestimmt haben.

In der **ersten Ebene**, dem Vergleich von Einleitung und Methodik, bestehen für den Schlüsselpunkt *DAS ZIEL DER INTERVENTION* drei mögliche Ursachen für eine inhaltliche Nichtübereinstimmung.

1. *Für eine formulierte Fragestellung in der Einleitung der Studie findet sich kein entsprechender definierter Endpunkt in der Methodik.*

2. *Für einen definierten Endpunkt in der Methodik der Studie findet sich in der Einleitung keine entsprechend formulierte Fragestellung.*

3. *Für eine formulierte Fragestellung in der Einleitung der Studie findet sich kein entsprechend definierter Endpunkt in der Methodik und gleichzeitig findet sich für einen definierten Endpunkt in der Methodik der Studie keine entsprechend formulierte Fragestellung in der Einleitung.*

Verglich man in der ersten Ebene den Schlüsselpunkt *DAS ZIEL DER INTERVENTION* zwischen der Einleitung und der Methodik, dann stimmten 24 der 53 Studien in diesem Schlüsselpunkt nicht überein.

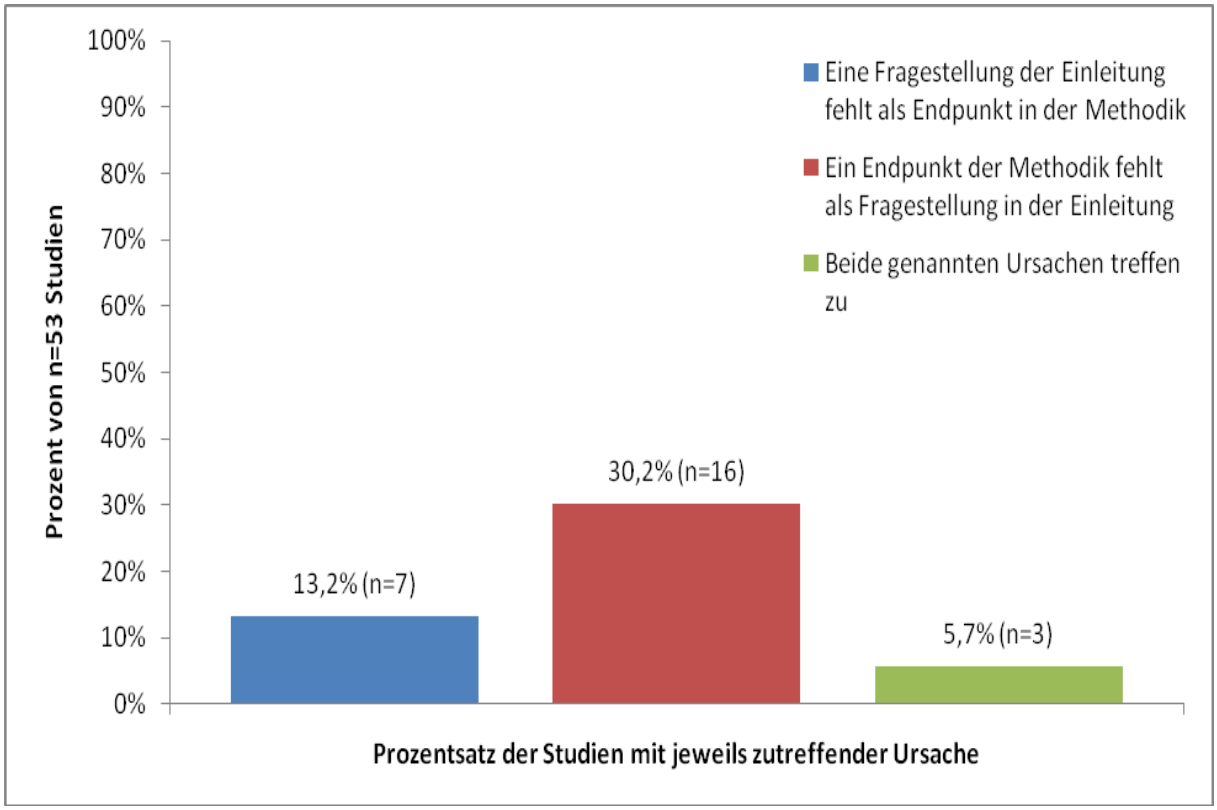

Abbildung 12: Analyse der Endpunkte – Ursachen der fehlenden Übereinstimmung im Schlüsselpunkt ZIEL DER INTERVENTION der ersten Ebene.

Sieben der 53 Studien stimmten in ihrem **Ergebnis** nicht überein, da in der Einleitung dieser Studien eine Fragestellung formuliert wurde, für die in der Methodik derselben Studie kein entsprechender Endpunkt definiert worden war. [80][95][100][107][110][123][130]

Insgesamt 16 der 53 untersuchten Studien definierten in der Methodik einen Endpunkt für den in der Einleitung keine entsprechende Fragestellung vorhanden war. [80][83][86][92][93] [101][103][109][110][111][114][115][122][127][128][130]

In drei der 53 Studien fehlte in der Methodik ein definierter Endpunkt für eine Fragestellung aus der Einleitung. Gleichzeitig definierte die Methodik dieser drei Studien einen Endpunkt, für den die Einleitung keine entsprechende Fragestellung formuliert hatte. [80][110][130]

In der **zweiten Ebene E2A**, dem Vergleich von Methodik und Diskussion, bestehen für den Schlüsselpunkt *DAS ERGEBNIS* fünf mögliche Ursachen für eine inhaltliche Nichtübereinstimmung.

1. *Für den primären Endpunkt aus der Methodik der Studie findet sich kein beobachtetes Ergebnis in der Diskussion.*

2. *Für einen sekundären Endpunkt aus der Methodik der Studie findet sich kein beobachtetes Ergebnis in der Diskussion.*

3. *Für einen zusätzlichen Endpunkt aus der Methodik der Studie findet sich kein beobachtetes Ergebnis in der Diskussion.*

4. *Für ein beobachtetes Ergebnis aus der Diskussion der Studie findet sich kein definierter Endpunkt in der Methodik.*

5. *Eine Zuordnung zu den bisher genannten Ursachen ist nicht möglich*

Verglich man in der zweiten Ebene E2A den Schlüsselpunkt *DAS ERGEBNIS* zwischen der Methodik und der Diskussion, dann stimmten 36 der untersuchten 53 Studien an diesem Schlüsselpunkt nicht überein.

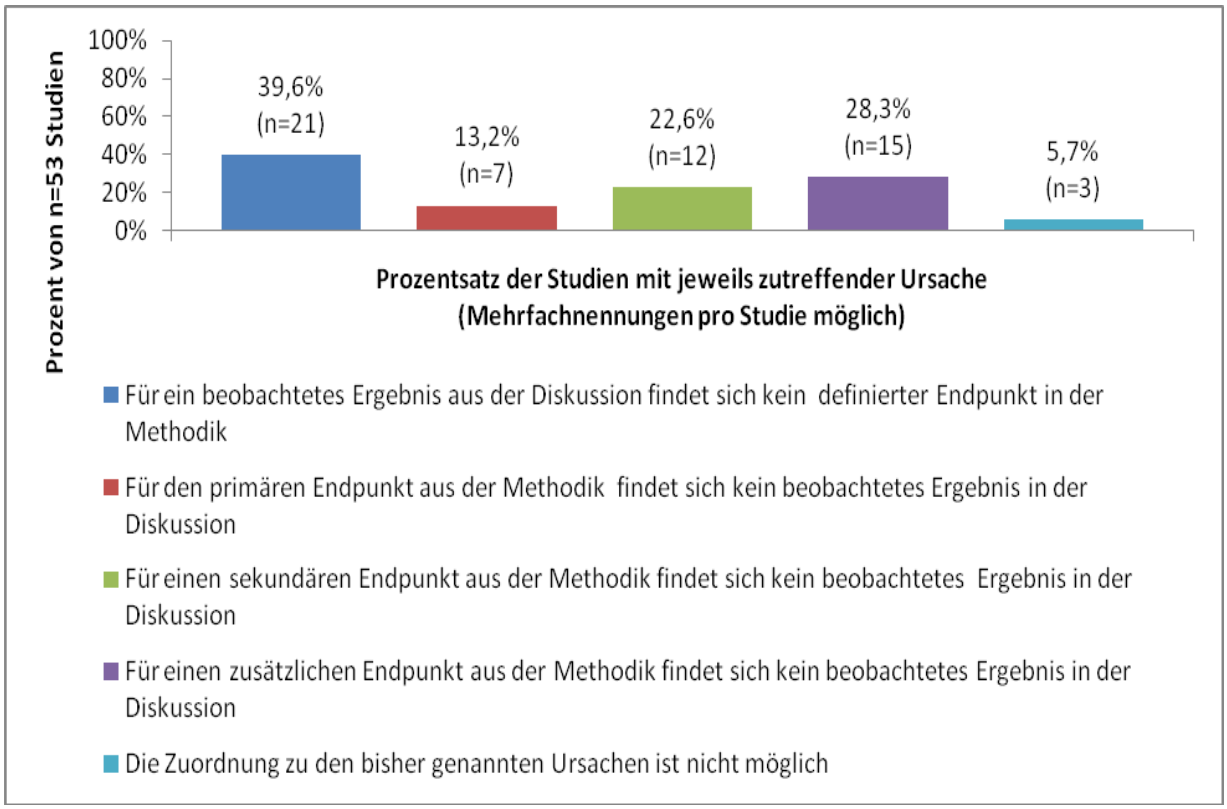

Abbildung 13: Analyse der Endpunkte – Ursachen der fehlenden Übereinstimmung im Schlüsselpunkt ERGEBNIS der zweiten Ebene E2A

In 21 Studien stimmte der Schlüsselpunkt „**Das Ergebnis**“ dieser Studien nicht überein, da in der Diskussion ein Ergebnis beobachtet worden war, für das die Methodik dieser Studie zuvor keinen Endpunkt definiert hatte. [79][81][85][86][87][93][95][98][100][101][102][107][109] [110][112][115][117][121][123][124][130]

Sieben Studien definierten in der Methodik zwar den primären Endpunkt, beobachteten dafür aber in der Diskussion derselben Studie kein Ergebnis. [86][100][104][117][124][129][130]

In 12 der 53 Studien wurde in der Methodik ein sekundärer Endpunkt definiert, ohne dass dafür in der Diskussion ein Ergebnis beobachtet worden wäre. [80][81][94][96][101] [102][104][112][117][126][127][129]

Insgesamt 15 Studien definierten in der Methodik zusätzliche Endpunkte, lieferten dafür allerdings in der Diskussion kein beobachtetes Ergebnis. [79][86][87][88][89][90][94][106] [109][110][111][115][120][129][131]

Drei Studien stimmten im Ergebnis aus anderen Gründen nicht überein.

In der **zweiten Ebene E2B**, dem Vergleich von Diskussion und Abstract, bestehen für den Schlüsselpunkt *DAS ERGEBNIS* ebenfalls fünf mögliche Ursachen für eine inhaltliche Nichtübereinstimmung.

1. *Das für den primären Endpunkt in der Diskussion der Studie beobachtete Ergebnis fehlt im Abstract.*

2. *Das für einen sekundären Endpunkt in der Diskussion der Studie beobachtete Ergebnis fehlt im Abstract.*

3. *Das für einen zusätzlichen Endpunkt in der Diskussion der Studie beobachtete Ergebnis fehlt im Abstract.*

4. *Für ein Ergebnis im Abstract der Studie findet sich kein beobachtetes Ergebnis in der Diskussion.*

5. *Eine Zuordnung zu den bisher genannten Ursachen ist nicht möglich.*

Verglich man in der zweiten Ebene E2B den Schlüsselpunkt *DAS ERGEBNIS* zwischen der Diskussion und dem Abstract, dann stimmten 46 der untersuchten 53 Studien an diesem Schlüsselpunkt nicht überein.

In 15 Studien stimmte der Schlüsselpunkt „**Das Ergebnis**“ nicht überein, da im Abstract der Studien ein Ergebnis stand, dass in der Diskussion derselben Studie so nicht beobachtet worden war. [80][86][87][89][91][97][107][110][112][115][117][123][124][130][131]

Zwei Studien hatten zwar für den primären Endpunkt dieser Studien in der Diskussion ein Ergebnis beobachtet, allerdings fehlte dieses Ergebnis im Abstract. [93][119]

17 Studien beobachteten in der Diskussion Ergebnisse für definierte sekundäre Endpunkte. Allerdings fehlten diese Ergebnisse im Abstract derselben Studien. [80][83][86][90][92][93][94][96][101][102][104][109][114[120][121][127][129]

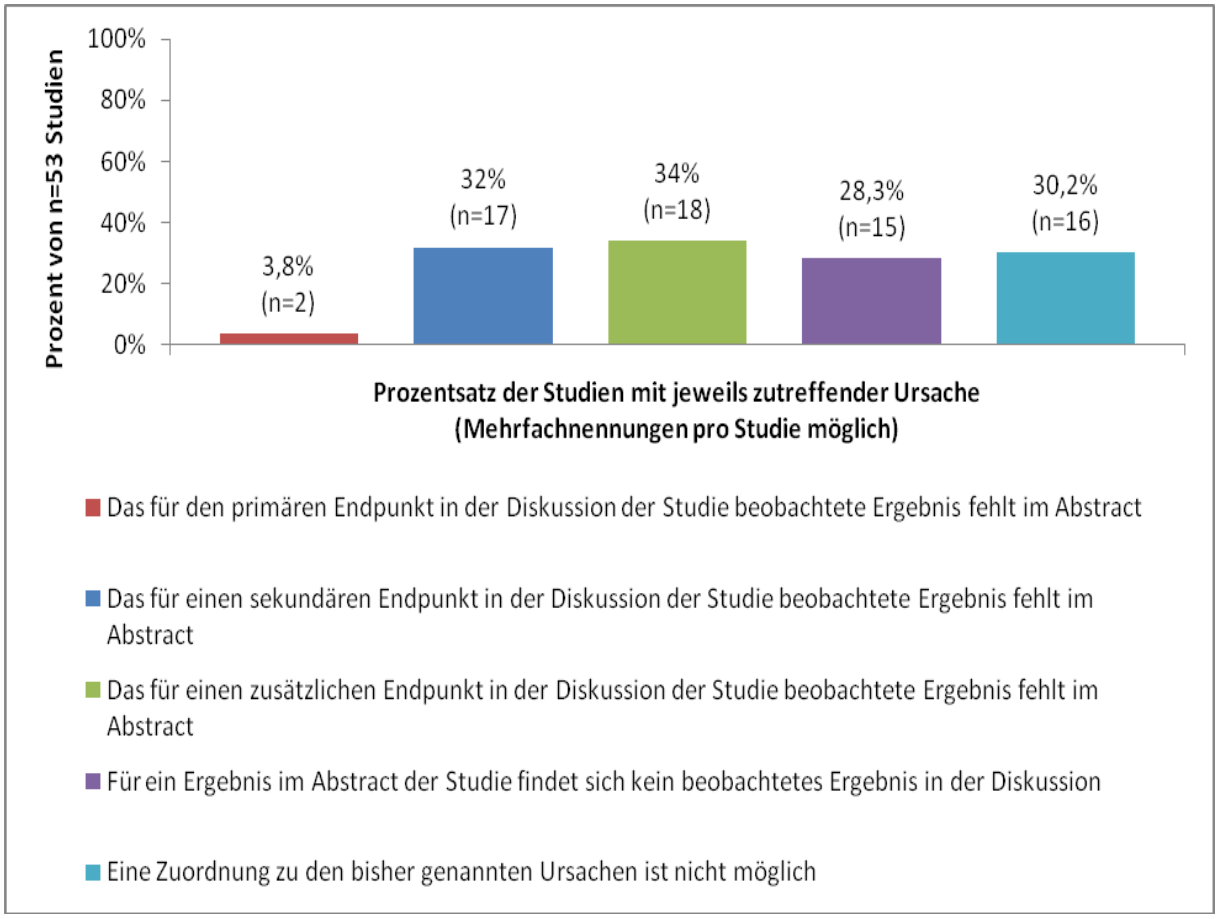

Abbildung 14: Analyse der Endpunkte – Ursachen der fehlenden Übereinstimmung im Schlüsselpunkt ERGEBNIS der zweiten Ebene E2B

18 Studien beobachteten in der Diskussion Ergebnisse für zusätzliche Endpunkte. Im Abstract der entsprechenden Studien fehlten diese Ergebnisse jedoch. [80][81][88][89][90][91][94][97][104][106][110][111][113][116][118][120][122][124]

16 Studien stimmten aus anderen Gründen nicht überein.

Andere Ursachen für eine fehlende Übereinstimmung am Schlüsselpunkt DAS ERGEBNIS in der zweiten Ebene:

Für insgesamt 19 Studien, die im Schlüsselpunkt DAS ERGEBNIS in der zweiten Ebene nicht übereingestimmt hatten, waren andere Ursachen, wie die oben genannten, ausschlaggebend für eine Nichtübereinstimmung. Diese werden auf der folgenden Seite in Tabelle 13 kurz dargestellt.

Tabelle 13: Weitere Ursachen für eine fehlende Übereinstimmung am Schlüsselpunkt DAS ERGEBNIS der zweiten Ebene

Weitere Ursachen für eine fehlende Übereinstimmung am Schlüsselpunkt DAS ERGEBNIS in der zweiten Ebene	*Die Anzahl der Studien, in der diese Ursache zu finden ist*	*Der prozentuale Anteil dieser Studien an der Gesamtzahl von 53 Studien*
In der Diskussion der Studie wurde ein Ergebnis beobachtet. Für dieses Ergebnis wurde aber kein Endpunkt definiert. Gleichzeitig fehlte dieses Ergebnis auch im Abstract. [81][85][86][87][95][98][100][102][107][117][121][124][130]	13	22,7 Prozent
Die bearbeitete klinische Studie definierte keinen primären Endpunkt. [112][125][126]	3	5,7 Prozent
Ergebnisse aus dem Abstract widersprechen teilweise beobachteten Ergebnissen aus der Diskussion. [85][129]	2	3,8 Prozent
Beobachtete Ergebnisse aus der Diskussion wurden mit wichtigen Einschränkungen versehen. Diese Einschränkungen fehlten im Abstract der Studie. [82][88]	2	3,8 Prozent
Im Abstract der Studie wurden Ergebnisse mit Einschränkungen versehen. In der Diskussion dieser Studie fehlten diese Einschränkungen bei den beobachteten Ergebnissen. [83]	1	1,9 Prozent
In der Diskussion der Studie wurde eine Empfehlung ausgesprochen, diese Empfehlung fehlte im Abstract. Diese Empfehlung war wichtig, bezog sich aber nicht auf einen definierten Endpunkt der Studie. [111]	1	1,9 Prozent

3.5 Das Ergebnis zur Auswertung der vierten Ebene

Die vierte Ebene vergleicht die Schlussfolgerung im Abstract mit dem primären Endpunkt einer Studie. Dabei stellt sie die Frage, ob sich die Schlussfolgerung im Abstract auf das beobachtete Ergebnis für den primären Endpunkt bezieht.

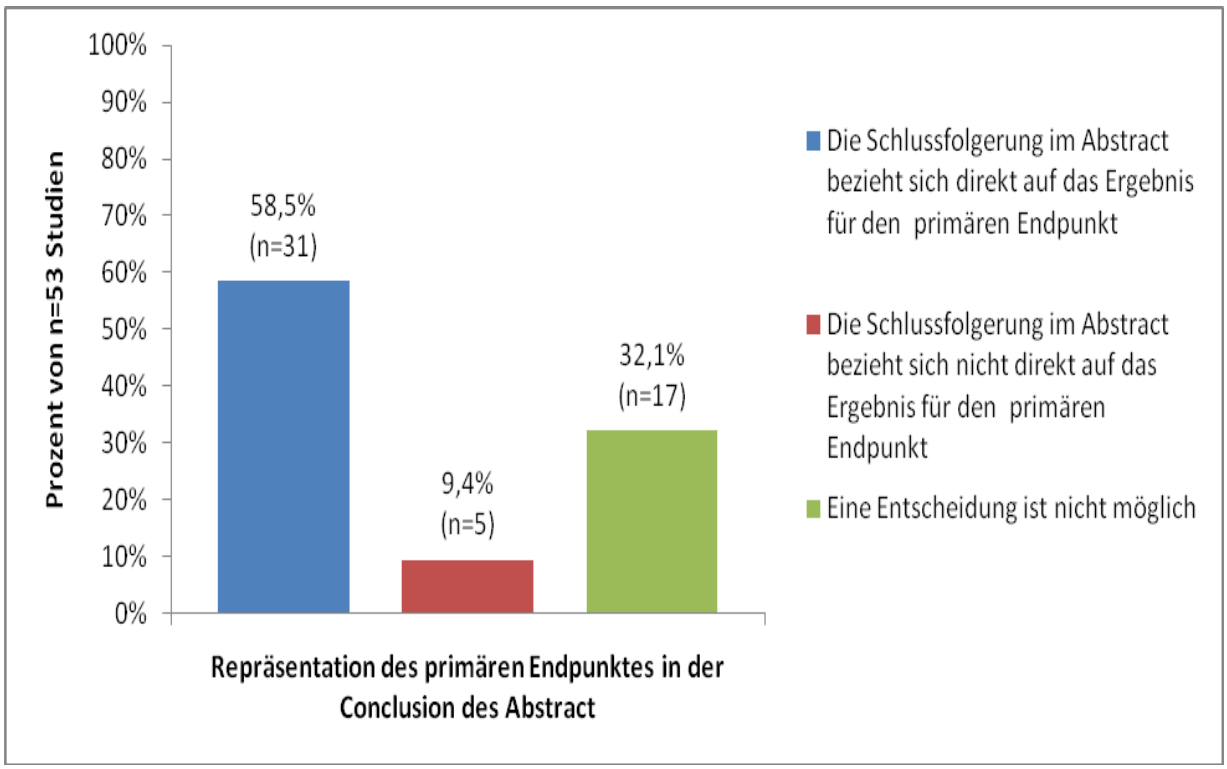

Abbildung 15: Der Stellenwert des primären Endpunktes in der Conclusion des Abstracts – Ergebnis der vierten Ebene

Verglich man das beobachtete Ergebnis für den primären Endpunkt einer Studie mit der Schlussfolgerung im Abstract, dann bezog sich diese Schlussfolgerung bei 31 Studien auf das Ergebnis des primären Endpunktes.

Fünf Studien zogen im Abstract eine Schlussfolgerung aus einem Ergebnis, dass nicht dem Ergebnis des primären Endpunktes entsprach. [99][100][112][125][126]

Bei 17 Studien war eine Entscheidung nicht möglich.

4. Diskussion

Die Evidenzbasierte Medizin ist der Versuch, die Diagnostik und Therapie von Erkrankungen auf die Grundlage harter wissenschaftlicher Daten zu stellen. Dazu bedient sich die Evidenzbasierte Medizin dem Werkzeug der randomisierten klinischen Studien. [2][3]
Durch den Vergleich von statistisch homogenen Gruppen soll eine Aussage getroffen werden, ob durch eine bestimmte Intervention ein Vorteil oder ein Nachteil durch diese Intervention für die jeweiligen Gruppen entsteht. Dieser Vorteil besteht im Idealfall in einer Verbesserung der Lebensqualität oder Lebensquantität. Auf den klinischen Alltag übertragen profitiert der individuelle Patient, dessen Charakteristika dem Studienkollektiv entsprechen, mit statistisch hinreichender Wahrscheinlichkeit von einer der klinisch geprüften Therapieoptionen. [5]
Da also auf der Grundlage der Ergebnisse klinischer Studien wichtige Entscheidungen zur alltäglichen Behandlung einer Vielzahl von Patienten getroffen werden, müssen an die Durchführung, Auswertung und Interpretation der Daten klinischer Studien höchste Qualitätsansprüche gestellt werden. [1][6] Das gilt neben ökonomischen Gründen nicht zuletzt auch aus ethischen Erwägungen.
Die Einführung neuer Therapieoptionen erfordert zumeist einen höheren finanziellen Aufwand als die Anwendung älterer Therapieverfahren. In Zeiten rasant gestiegener Gesundheitskosten befinden sich die Entscheidungsträger im klinischen Alltag also in einem zunehmenden Spannungsfeld zwischen dem medizinisch Erforderlichen und dem finanziell Machbaren. [73]
Aus ethischen Gründen muss man sich zweier Dinge vergegenwärtigen. Zum einen bedeutet die Teilnahme an einer klinischen Studie ein nicht zu unterschätzendes Risiko für jeden einzelnen Patienten. Schließlich besteht die Möglichkeit, dass die geprüften Therapieoptionen mehr Schaden bereiten als Nutzen. Dennoch erfolgt die Teilnahme an klinischen Studien zumeist in dem festen Glauben, dadurch einen wichtigen Beitrag zum medizinisch-wissenschaftlichen Fortschritt zu leisten. [6]
Zum anderen kann durch die Fehlinterpretation der Daten klinischer Studien und der daraus abgeleiteten Therapieanwendungen ein Schaden entstehen, der unter Umständen bis zum Tod des einzelnen Patienten reicht. Man erinnere an dieser Stelle nur an den Einsatz von antiarhythmisch wirksamen Medikamenten bei speziellen Formen ventrikulärer Rhythmusstörungen in den 80er und frühen 90er Jahren. Durch die Publikation der *CAST II* Studie 1992 konnte gezeigt werden, dass der Einsatz dieser Medikamente zu einer Übersterblichkeit in der Gruppe der behandelten Patienten geführt hatte. [74] Es starben also

über Jahrzehnte viele Patienten an den Folgen einer Therapie von der sie eigentlich hätten profitieren sollen.

Dass nicht alle publizierten klinischen Studien das erforderliche Höchstmaß an Qualität in der Durchführung, Auswertung und Interpretation der Daten dieser Studien erfüllen, ist in der wissenschaftlichen Gemeinschaft ein intensiv diskutiertes Problem. [9][12][17][40]

Ein Blick in unterschiedliche Publikationen zu diesem Thema fördert eine große Anzahl bekannter Fehlerquellen zu Tage, die zu einer Falschinterpretation der Daten klinischer Studien führen können.

Eine der bekanntesten Fehlerquellen ist der *„publication bias“*. Durch die selektive Publikation von Studien mit einem positiven Ergebnis für eine Therapieoption wird erreicht, dass die klinischen Effekte dieser Therapieoption oft überschätzt werden. [57][58][38] Studien hingegen, die keinen Unterschied durch die Anwendung dieser Therapieoption fanden [52][53] oder die, im Gegenteil, einen negativen Effekt durch die Anwendung beobachteten, werden häufig nicht publiziert. [54][60]

Eine weitere mögliche Fehlerquelle entsteht durch den vorzeitigen Stopp klinischer Studien nach einer geplanten Zwischenauswertung der bisherigen Daten. Der Stopp erfolgt in den meisten Fällen, da ein deutlicher signifikanter klinischer Vorteil für die eine Therapieoption im Vergleich zur anderen gefunden wurde. Dieser Vorteil wird aber oft überschätzt. [10] Denn es konnte in vielen Fällen gezeigt werden, dass sich anfängliche signifikante Therapieeffekte im Langzeitverlauf einer Studie relativierten, sodass die klinische Bedeutung dieser Effekte durch einen vorzeitigen Stopp der Studie überschätzt worden wäre. [41]

Eine weitere sehr bedeutende Fehlerquelle stellen Protokollverletzungen dar. Das bedeutet, dass trotz fester Zuteilung eines Studienteilnehmers zu einem Interventionsarm der Studie dieser Studienteilnehmer eine andere Intervention erhielt als nach dem Studienprotokoll geplant war. Protokollverletzungen sind sehr häufig in klinischen Studien und bei der meist großen Zahl an Studienteilnehmern auch kaum zu verhindern. [26] Entsprechend gibt es unterschiedliche Ansätze zur Auswertung dieser Daten. Man kann die betroffenen Studienteilnehmer weiterhin der ursprünglich zugeteilten Gruppe zurechnen, das entspricht dem so genannten *„per protocol“* Ansatz. Man kann die Studienteilnehmer aber auch, der erhaltenen Therapie entsprechend, der anderen Gruppe zurechnen. Das entspräche der *„intention to treat“* Analyse der Daten. [32] Es konnte gezeigt werden, dass die Datenauswertung einer klinischen Studie je nach gewähltem Ansatz zu einem unterschiedlichen Ergebnis führen kann. [29][30] Dabei beschreibt die *„intention to treat“* Analyse der Daten am ehesten den wahren klinischen Effekt einer Intervention. [32]

Eine andere Fehlerquelle sind die Schlussfolgerungen aus den Ergebnissen klinischer Studien für nachträglich definierte Subgruppen. Subgruppen werden gebildet, indem nach der Datenauswertung Studienteilnehmer mit bestimmten Charakteristika aus der ursprünglichen Studiengruppe herausgenommen werden. Diese werden dann in Subgruppen eingeteilt. Anschließend erfolgt eine erneute Auswertung der Daten für diese Subgruppen. Dadurch ist es möglich, zumindest für einen Teil der ursprünglichen Studiengruppe, statistisch signifikante Effekte zu finden, die vorher nicht gefunden werden konnten. [37][51] Die Bildung von Subgruppen ist im ursprünglichen Studienprotokoll nicht definiert, da sie erst nach der Datenauswertung gebildet werden. Trotzdem ist die Auswertung nachträglich definierter Subgruppen durch statistische Berechnungen möglich. Allerdings muss die Bildung von Subgruppen klaren Regeln folgen um falsch positive Ergebnisse zu vermeiden. Sonst besteht die Gefahr, dass die wahren klinischen Effekte eines signifikanten Ergebnisses überschätzt werden [48][49]

Alle diese genannten Fehlerquellen haben gemeinsam, dass sie sich entweder auf die Publikation klinischer Studien beziehen, oder die Datenauswertung klinischer Studien betreffen. *Chan et al* publizierten 2004 im *JAMA* einen Artikel in dem sie eine weitere Problematik erkannten.

Sie verglichen die Studienprotokolle mit den publizierten Originalarbeiten zu diesen Studien und stellten fest, dass diese häufig nicht nur unvollständig waren sondern auch falsch und mit dem Studienprotokoll nicht konsistent. Was sie beschrieben war eine selektive Veröffentlichung der Ergebnisse randomisierter klinischer Studien. Sie mussten feststellen, dass viele Originalarbeiten für einen Teil der ursprünglich im Studienprotokoll definierten Endpunkte keine Ergebnisse geliefert hatten. [12] Wenn also die Protokolle klinischer Studien zu einem Teil nicht mit dem Inhalt der Originalarbeiten konsistent sind, dann stellt sich die Frage wie konsistent der Inhalt der Originalarbeit selbst ist. Diese Frage beantworten die Ergebnisse der ICCT.

Würde man voraussetzen, dass randomisierte klinische Studien in ihrem Inhalt konsistent sind, dann könnte man daraus die Schlussfolgerung ziehen, dass es möglich ist eine Stelle im Text dieser Studien zu finden, die alle relevanten Informationen enthält. Das würde auch bedeuten, dass es möglich ist den Inhalt der Studie komplett zu erfassen ohne dafür die gesamte Arbeit lesen zu müssen. Wären die Studien in ihrem Inhalt aber nicht konsistent so könnte man zum einen beschreiben, an welchen Stellen im Text die Inhalte nicht konsistent sind und zum anderen, welche Gründe dafür ausschlaggebend sind. Daraus wiederum könnten Schlussfolgerungen gezogen werden, auf welche Art und Weise sich die Berichterstattung

randomisierter klinischer Studien verbessern ließe, um mit hinreichender Sicherheit alle relevanten Informationen dieser Studie zu erhalten.

Die ICCT untersuchte die Konsistenz randomisierter klinischer Studien, indem sie den Inhalt verschiedener Textstellen miteinander verglich. Dazu wurden alle Informationen zu definierten Schlüsselpunkten in diesen Textstellen gesammelt. Stimmten diese Informationen überein, dann waren die Textstellen identisch. Die ICCT richtete ihr Augenmerk auf drei zentrale Punkte jeder klinischen Studie. Die Studiengruppe, die Intervention an dieser Studiengruppe und die beobachteten Ergebnisse aus dieser Intervention. [2] Zu diesen Punkten wurden die Informationen in den Textstellen gesammelt. Zum Vergleich wurden die Textstellen Titel, Einleitung, Methodik, Diskussion und Abstract der jeweiligen Studie herangezogen.
Nach Durchsicht von 53 randomisierten klinischen Studien aus drei verschiedenen Journals stand fest, dass keine dieser 53 Studien durchgehend inhaltlich konsistent war. Es fehlten relevante Informationen entweder zur Studiengruppe, zur Intervention oder den beobachteten Ergebnissen in einer der zu vergleichenden Textstellen.
Daraus kann der Schluss gezogen werden, dass es nicht möglich ist eine bestimmte Textstelle in einer Studie zu finden, die alle relevanten Informationen dieser Studie enthält. Alle relevanten Informationen erhält man nur dann, wenn man die komplette Studie gelesen hat.

Dass keine der untersuchten Studien in ihrem Inhalt durchgehend konsistent war, bedarf jedoch unbedingt einer differenzierteren Betrachtung. Die ICCT betrachtete die Konsistenz einer randomisierten klinischen Studie auf drei verschiedenen Ebenen.
In der ersten Ebene prüfte sie die inhaltliche Konsistenz der geplanten Intervention einer klinischen Studie im Vergleich mit der tatsächlich durchgeführten Intervention dieser Studie. Dazu verglich sie die Einleitung der Studie mit der Methodik dieser Studie. Sie sammelte aus beiden Textstellen alle verfügbaren Informationen zur Studiengruppe, der Intervention und dem Ziel der Intervention und verglich diese innerhalb der beiden Textstellen.
In der zweiten Ebene sammelte die ICCT alle verfügbaren Informationen zur Studiengruppe, den definierten Endpunkten und den beobachteten Ergebnissen. Sie verglich die Informationen zur Studiengruppe, den definierten Endpunkten und den beobachteten Ergebnissen zwischen der Methodik und der Diskussion. Anschließend verglich sie die Informationen über die Studiengruppe und die beobachteten Ergebnisse in der Diskussion und im Abstract.

In der dritten Ebene prüfte die ICCT, ob der Studientitel die Erwartungen an die Intervention und die Fragestellung der betreffenden Studie tatsächlich erfüllte. Dazu verglich sie die Intervention und die Fragestellung im Titel mit diesen Punkten in der Methodik und der Diskussion.

Als erfreuliches Ergebnis konnte die ICCT zeigen, dass der Titel einer Studie die Erwartungen an die Intervention und die Fragestellung dieser Studie in der überwiegenden Zahl der untersuchten Studien erfüllte. Enthielt der Studientitel Informationen über die Intervention der Studie, so stimmten diese Informationen bei 49 der untersuchten 53 Studien überein. Da vier dieser Studien keine Informationen über die Intervention enthielten, entsprach die Intervention im Titel in 100 Prozent der untersuchten Studien der tatsächlich durchgeführten Intervention dieser Studien. Stellte der Titel einer Studie eine Fragestellung an die Studie, so fand sich bei 42 der untersuchten 53 Studien für diese Fragestellung auch ein beobachtetes Ergebnis. Lediglich bei vier Studien war eine Entscheidung nicht möglich, während sieben Studien keine Fragestellung im Titel formulierten. Es fand sich also keine Studie, welche die Fragestellung im Titel nicht beantwortete.

Dieses positive Bild ändert sich jedoch wenn man die Ergebnisse für die beiden anderen Ebenen betrachtet. Verglich man die Einleitung der Studie mit der Methodik dieser Studie, so stimmte deren Inhalt in 26 der untersuchten 53 Studien nicht überein. Die ICCT definierte die Einleitung als die Textstelle der geplanten Intervention einer Studie. Die Textstelle beschrieb eine Studiengruppe an der eine bestimmte Intervention, mit einer Fragestellung an diese Intervention verknüpft, durchgeführt werden sollte. Die geplante Intervention entsprach damit bei 49,1 Prozent aller Studien nicht der durchgeführten Intervention aus der Methodik. Das bedarf einer genaueren Differenzierung.

Zur Überprüfung der Konsistenz der beiden Textstellen Einleitung und Methodik wurden Informationen zu drei Schlüsselpunkten gesammelt: Der Studiengruppe, der Intervention und dem Ziel der Intervention. Verglich man die Studiengruppen aus der Einleitung und der Methodik, so stimmten diese bei 40 der untersuchten 53 Studien überein. Lediglich in drei Studien stimmten sie nicht überein. Der Vergleich der Intervention in den beiden Textstellen ergab eine Übereinstimmung bei 32 der untersuchten 53 Studien. Allerdings unterschieden sich die Informationen über die Intervention schon definitiv bei sechs der untersuchten 53 Studien. Diese Zahlen ändern sich deutlich beim letzten der drei Schlüsselpunkte, dem Ziel der Intervention. Verglich man die Informationen über die Ziele der Intervention in der Einleitung mit den definierten Endpunkten der Methodik, so stimmten hier nur 13 der

untersuchten 53 Studien überein. Bei 24 Studien stimmten die Ziele der Intervention nicht mit den definierten Endpunkten überein.

Die fehlende Konsistenz zwischen der Einleitung und der Methodik beruht also zum großen Teil auf dem Umstand, dass sich die geplanten Fragestellungen und die tatsächlich definierten Endpunkte in den Studien unterscheiden.

In der zweiten Ebene zeigt sich ein ähnliches Bild. Hier ergab der Vergleich der Textstellen Methodik und Diskussion, dass 36 der untersuchten 53 Studien nicht übereingestimmt hatten. Dieses Ergebnis basierte zum einen auf dem Vergleich der Studiengruppe der Methodik mit der Studiengruppe der Diskussion und zum anderen auf dem Vergleich der definierten Endpunkte der Methodik mit den beobachteten Ergebnissen der Diskussion. Die Studiengruppe stimmte dabei in 33 der 53 untersuchten Studien überein. Sie unterschied sich definitiv bei 15 Studien. Das entsprach einer Übereinstimmung von immerhin 62,3 Prozent. Der Vergleich der definierten Endpunkte der Methodik mit den beobachteten Ergebnissen der Diskussion ergab jedoch eine völlig andere Verteilung. Bei diesem Vergleich stimmten noch 11 der 53 Studien inhaltlich komplett überein. Insgesamt 36 Studien unterschieden sich definitiv. Das entsprach nur noch einer Übereinstimmung von 20,8 Prozent. Somit beruhte die fehlende Konsistenz zwischen der Methodik und der Diskussion zumeist darauf, dass sich die definierte Endpunkte und die beobachteten Ergebnisse unterschieden.

Für die Textstellen Diskussion und Abstract findet sich fast das gleiche Bild. In 46 der untersuchten 53 Studien stimmten die Schlüsselpunkte in beiden Textstellen nicht überein. Das entspricht 86,8 Prozent aller untersuchter Studien.

Auch hier erfolgte ein Vergleich der Studiengruppe zwischen der Diskussion und dem Abstract. Ebenso wurden die beobachteten Ergebnisse der Diskussion mit den Ergebnissen aus dem Abstract verglichen. Die Studiengruppe stimmte bei 35 der untersuchten 53 Studien überein. Bei neun Studien stimmte sie definitiv nicht überein. Das entsprach einer Übereinstimmung von 66 Prozent. Verglich man aber die beobachteten Ergebnisse der Diskussion mit den Ergebnissen im Abstract, dann stimmten nur noch fünf Studien überein. Das entspricht 9,4 Prozent. Insgesamt 46 Studien, das entspricht 86,8 Prozent aller untersuchter Studien, stimmten nicht überein. Die fehlende Konsistenz zwischen der Diskussion und dem Abstract beruhte also bei jeder untersuchten Studie darauf, dass sich die Ergebnisse in diesen beiden Textstellen unterschieden.

Diese Zahlen sind beeindruckend. Sie lassen zu dem Schluss kommen, dass bei der Präsentation der Ergebnisse randomisierter klinischer Studien wichtige Informationen

verloren gehen. Dieser Informationsverlust kann zu einer Fehlinterpretation der Ergebnisse klinischer Studien führen.

Scherer et al verglichen 2007 die Kongresszusammenfassungen klinischer Studien mit den später publizierten Originalarbeiten. Dabei fanden sie heraus, dass nur 62 Prozent dieser Abstracts als vollständige Originalarbeiten publiziert wurden. [45] Ähnliche Ergebnisse fand eine Gruppe um *Toma et al*. Zusätzlich stellten sie fest, dass sich die Ergebnisse in den Abstracts häufig von den Ergebnissen in der Originalarbeit unterschieden. [43]

Die ICCT konnte zeigen, dass der Abstract klinischer Studien bei 46 der untersuchten 53 Studien nicht mit der Diskussion übereinstimmte. Das bestätigt die oben genannten Zahlen. Wenn aber bei der Präsentation der Ergebnisse klinischer Studien in einer großen Anzahl der Studien Informationen verloren gehen, dann führt dies zwangsläufig zu der Frage, ob diesen Ergebnissen dann noch zu trauen ist. Diese Frage kann die ICCT nicht mit Sicherheit beantworten. *Soares et al* publizierten 2004 im *British Medical Journal* eine Arbeit in der sie feststellten, dass die unvollständige Präsentation von Ergebnissen klinischer Studien nicht zwangsläufig zu falschen Schlussfolgerungen führen muss. [34] Aber ein Blick auf die Gründe für die fehlende Konsistenz der Ergebnisse lohnt sich dennoch. Denn er fördert interessante Aspekte zu Tage.

Verglich man die geplante Fragestellung der Einleitung mit den definierten Endpunkten der Methodik, dann stimmten diese bei 24 der untersuchten 53 Studien nicht überein. Dabei beschrieben sieben Studien eine Fragestellung in der Einleitung, für die sie später keinen Endpunkt definierten. 16 Studien definierten Endpunkte ohne diese bereits in der Einleitung beschrieben zu haben.

Verglich man nun die endgültig definierten Endpunkte mit den beobachteten Ergebnissen, dann stimmten diese bei 36 der untersuchten 53 Studien nicht überein. Den größten Anteil hatte daran interessanterweise der Fall, dass Studien Ergebnisse beobachtet hatten ohne dafür einen Endpunkt zu definiert. Das traf für 21 Studien zu. Immerhin entspricht das 39,4 Prozent aller Studien. 15 Studien definierten zusätzliche Endpunkte in der Methodik ohne dafür ein entsprechendes Ergebnis zu beobachten. In 12 Studien geschah dieses für sekundäre Endpunkte. Bei immerhin sieben Studien, das entspricht 13,2 Prozent, wurden primäre Endpunkte definiert, ohne hierfür Ergebnisse zu liefern. Eine beachtliche Zahl, führt man sich vor Augen, dass der primäre Endpunkt die Hauptfragestellung einer Studie ist. Alle statistisch relevanten Charakteristika werden exakt für diesen Endpunkt berechnet. [7][14][15]

Verglich man zum Schluss die beobachteten Ergebnisse der Diskussion mit den Ergebnissen im Abstract, dann stimmten diese in 46 der untersuchten 53 Studien nicht überein. Das hatte

unterschiedliche Gründe. In 18 Studien fehlten beobachtete Ergebnisse für zusätzliche Endpunkte im Abstract. In 17 Studien fehlten diese Ergebnisse für sekundäre Endpunkte. Interessanterweise präsentierten 15 Studien ein Ergebnis im Abstract, ohne dass dieses Ergebnis in der Diskussion beobachtet worden wäre. 13 Studien beobachteten Ergebnisse allein in der Diskussion ohne hierfür einen Endpunkt zu definieren oder diese Ergebnisse im Abstract zu präsentieren. Immerhin noch zwei Studien beobachteten in der Diskussion ein Ergebnis für den primären Endpunkt, ohne dass sich dieses Ergebnis im Abstract wieder gefunden hätte.

Die Ergebnisse in randomisierten klinischen Studien kann man folglich nicht gerade als stringent bezeichnen. Endpunkte werden definiert ohne Ergebnisse zu liefern, Ergebnisse werden präsentiert ohne entsprechende Endpunkte. *Hrobjartsson et al* veröffentlichten 2005 im *Ugeskr Laeger* eine interessante Arbeit die diesen Befund unterstreicht. Sie stellten fest, dass bei 50 Prozent der von Ihnen geprüften Studien die Ergebnisse zu den definierten Endpunkten nicht vollständig waren. Dabei fanden sich häufiger Ergebnisse für die Endpunkte, deren Ergebnis statistisch signifikant war. 62 Prozent der Studien hatten darüber hinaus im Verlauf den primären Endpunkt geändert. [55] *Wood et al* beschrieben ähnliches bereits 2004 im *Clinical Trials*. Sie stellten fest, dass in 89 Prozent der von ihnen untersuchten Studien Ergebnisse für die definierten Endpunkte fehlten. [56]

Das lässt eine gewisse Willkürlichkeit im Umgang mit den Ergebnissen randomisierter klinischer Studien erkennen. Las man die 53 Originalarbeiten der ICCT so fiel auf, dass eine große Zahl an Studien eine Unmenge an verschiedenen Endpunkten definierte. Es kam nicht selten vor, dass selbst primäre Endpunkte mehrfach in Studien vorhanden waren.

Zuweilen erinnerte das beinahe an eine Entenjagd mit Schrottkugeln. Je mehr Kugeln durch einen Schuss abgefeuert werden, umso größer ist die Wahrscheinlichkeit wenigstens eine Ente zu treffen. Wie gut gezielt wird ist dabei letzlich egal. Übertragen auf randomisierte klinische Studien erhöht sich also die Wahrscheinlichkeit für ein signifikantes Ergebnis, je mehr Endpunkte definiert werden. [75]

Die klinische Relevanz von Ergebnissen für zusätzliche Endpunkte mag nicht allzu bedeutend sein. Gleiches gilt zuweilen für die Ergebnisse von sekundären Endpunkten. [34] Wenn aber sieben Studien in der ICCT einen primären Endpunkt definieren und für diesen kein Ergebnis präsentieren und gleichzeitig 21 Studien Ergebnisse beobachten ohne Endpunkte zu definieren, dann sollte durchaus die Frage erlaubt sein, inwieweit den Ergebnissen dieser Studien dann noch zu trauen ist. Für eine Antwort sollte man noch ein weiteres Ergebnis der ICCT betrachten.

Die ICCT hatte untersucht, wie oft sich die Schlussfolgerung im Abstract auf das Ergebnis für den primären Endpunkt bezieht. Dies war in 31 der untersuchten 53 Studien der Fall. In 17 Studien war einen Entscheidung nicht möglich. Fünf Studien bezogen sich in der Schlussfolgerung im Abstract nicht auf den primären Endpunkt. Das waren 9,4 Prozent aller Studien.

Man sollte sich daran erinnern, dass in einer klinischen Studie nur der primäre Endpunkt die notwendige Power besitzt, einen signifikanten Unterschied zwischen den Studiengruppen zu finden, der mit einer hinreichenden Wahrscheinlichkeit als wahr zu betrachten ist. [7][14][15] Daher ist es nicht zulässig eine gültige Schlussfolgerung aus einer Studie zu ziehen, wenn sich diese Schlussfolgerung nicht auf den primären Endpunkt bezieht. Aus dem gleichen Grund sollte man das Ergebnis einer Studie anzweifeln, die für den primären Endpunkt kein Ergebnis liefert.

Ähnliches gilt, wenn für Ergebnisse einer Studie keine Endpunkte definiert wurden. Beobachtet eine Studie Ergebnisse in der Diskussion, ohne dafür im Voraus Endpunkte definiert zu haben, dann können aus diesen Ergebnissen keine Schlussfolgerungen gezogen werden. Die Hypothesen einer klinischen Studie müssen vor deren Durchführung erstellt werden. Auf Grundlage dieser Hypothesen werden die Endpunkte der Studie erstellt. Nur für Ergebnisse aus vordefinierten Endpunkten kann dann eine Schlussfolgerung getroffen werden. Ergeben sich durch die Auswertung der Daten weitere Ergebnisse, so dürfen diese zwar präsentiert werden, es muss jedoch deutlich gemacht werden, dass für diese Ergebnisse keine statistische Gültigkeit besteht, da sie erst nachträglich gefunden wurden. [2][3]

Die ICCT konnte zeigen, dass randomisierte klinische Studien in ihrem Inhalt nicht konsistent sind. In der überwiegenden Anzahl der untersuchten Studien war dies auf eine Inkonsistenz der Ergebnisse zurückzuführen. Die Gründe dafür wurden ausführlich dargelegt und decken sich in großen Teilen mit dem bisherigen Stand der wissenschaftlichen Forschung. [12][55][56] Dass die Inkonsistenz der Ergebnisse nicht unbedingt einen Rückschluss auf die Gültigkeit dieser Ergebnisse zulässt wurde ebenfalls erläutert. [34] Dennoch konnte die Schlussfolgerung gezogen werden, dass zumindest für einen Teil der Studien die Gültigkeit der gefundenen Ergebnisse angezweifelt werden muss.

Über die Gründe, die zum Ergebniss der ICCT geführt haben, wurde bisher noch keine Aussage gemacht. Dabei ist die Frage durchaus berechtig, warum so viele Studien nur unvollständig über die Ergebnisse berichten. Eine Antwort auf diese Frage auf der Datenbasis der ICCT zu geben ist nicht möglich. Zieht man jedoch einige interessante Publikationen der letzten Jahre zu Hilfe, so findet sich durchaus ein Erklärungsansatz.

Lee et al veröffentlichten 2008 im *PLoS Medicine* eine Übersichtsarbeit in der sie feststellten, dass über die Hälfte der Zulassungsstudien für Medikamente auch fünf Jahre nach deren Zulassung durch die FDA noch nicht publiziert worden waren. Studien mit positiven Ergebnissen wurden dabei häufiger veröffentlicht. [59] *Turner et al* erbrachten den Nachweis, dass durch die selektive Veröffentlichung der Studien zum Wirksamkeitsnachweis von Antidepressiva deren Wirksamkeit bei weitem überschätzt wurde. [38] *Melander et al* beschrieben 2003 im *British Medical Journal* die Häufigkeit selektiver Veröffentlichungen von Studien durch die pharmazeutische Industrie und die draus resultierende Überschätzung des klinischen Effektes der geprüften Substanzen. [58]

Die Entwicklung und Erforschung neuer Medikamente ist sehr kostenintensiv. So beträgt die durchschnittliche Entwicklungszeit eines neuen Medikaments zwischen zehn und zwölf Jahren. Die Kosten können dabei leicht mehrere Milliarden Euro betragen. [76] Die Zulassung eines neuen Medikaments ist nur nach sehr strengen Kriterien möglich. Ohne die Daten aus großen Zulassungsstudien können diese Kriterien nicht erfüllt werden. [77] Erbringen diese Zulassungsstudien keinen Wirksamkeitsnachweis, dann kann dies massive finanzielle Verluste zur Folge haben. Die Kostenerstattung durch die gesetzliche Krankenversicherung in Deutschland ist beispielsweise an den Wirksamkeitsnachweis gebunden. [78] *Gotzsche et al* legten 2006 im *Journal of American Mediacl Association* dar, wie groß der Einfluss der pharmazeutischen Industrie auf die Veröffentlichung klinischer Studien ist. So stellten sie fest, dass in über 50 Prozent der Studienprotokolle, die sie untersucht hatten, dem Finanzier der Studie die Möglichkeit eingeräumt wurde, die Publikation der Studie zu verhindern. [36]

Ohne die finanzielle Unterstützung durch entsprechende Geldgeber ist die Durchführung einer klinischen Studie kaum möglich. Das führt dazu, dass ein Großteil der klinischen Studien aus Geldern der pharmazeutischen Industrie finanziert wird. Forschung ohne finanzielle Abhängigkeiten ist kaum mehr möglich, die Einflussmöglichkeiten von Außen auf die medizinische Forschung wachsen. [35]

Selbstverständlich ist es nicht möglich, aus diesen Überlegungen den Schluss zu ziehen, dass die Ergebnisse klinischer Studien bewusst unvollständig sind. Diesen Schluss lassen die Daten der ICCT auch gar nicht zu. Der Blick in die Literatur zeigt aber, wie abhängig die medizinische Forschung von finanziellen Zuwendungen geworden ist. Wie unabhängig sie unter diesen Bedingungen noch arbeiten kann, sei dahingestellt. Dass durch die selektive Veröffentlichung der Ergebnisse klinischer Studien die Effekte einer Therapie oft überschätzt werden, [38][57][58][61] ist bekannt. Dass daraus wiederum ein finanzieller Vorteil entstehen

kann, ist durchaus möglich. Schließlich wird die Anwendung einer Therapie im klinischen Alltag umso wahrscheinlicher, je besser die Therapie wirkt. Wie gut sie wirkt, dass entscheiden im Endeffekt die Ergebnisse klinischer Studien.

Aus den Ergebnissen der ICCT müssen Konsequenzen gezogen werden. Die Inkonsistenz in der Präsentation der Ergebnisse klinischer Studien ist so nicht akzeptabel. Gleiches gilt für die gefundene Inkonsistenz der Intervention und der Studiengruppe in klinischen Studien. Zwar erreichte die Inkonsistenz der Intervention und der Studiengruppe nicht das Ausmaß der Inkonsistenz der Ergebnisse, sie kann aber dennoch nicht ignoriert werden. Als Konsequenz muss man aus den Ergebnissen der ICCT folgern, dass die Berichterstattung klinischer Studien insgesamt verbessert werden muss.

Es gab bereits einige sehr erfolgreiche Ansätze um die Berichterstattung klinischer Studien zu verbessern. So publizierten *Moher et al* 2001 eine revidierte Version des bereits durch *Begg et al* 1996 vorgestellten CONSORT Statement [4][67] zur Verbesserung der Qualität in der Berichterstattung randomisierter klinischer Studien. Es wurde eine Checkliste mit 22 Punkten gefordert, die in jeder klinischen Studie vorhanden sein sollten. Zu diesen Punkten gehörten beispielsweise die Gruppenspezifikationen, die definierten Endpunkte oder das Verblindungsmuster für die Gruppenzuteilung. Es wurden Regeln aufgestellt, wie über diese Punkte zu berichten ist. [4][33] Durch das CONSORT Statement konnte die Validität und somit letztendlich die Beweiskraft klinischer Studien entscheidend verbessert werden. [63] [64]

Der Problematik des „publication bias“ begegnete man, indem das 2005 publizierte Ottawa Statement die Registrierung und vollständig Veröffentlichung des Studienprotokolls jeder klinischen Studie forderte. [39] Dadurch wurde es möglich, zum einen die Studienprotokolle mit den Originalarbeiten zu vergleichen und zum anderen überhaupt zu bemerken, dass Studien initiiert wurden, deren Ergebnisse bisher noch nicht veröffentlicht wurden. [66] Durch den Zugang zu den Protokollen verbessert sich auch die Transparenz der gefundenen Ergebnisse, da nun ein direkter Vergleich mit den definierten Endpunkten möglich war. Das wiederum erhöhte letztlich auch die Validität der Studien. [65]
Die ICCT fordert, dass die Berichterstattung klinischer Studien weiter verbessert werden muss. Denn was bedeutet fehlende inhaltliche Konsistenz? Sie bedeutet im Endeffekt nichts anderes als den Verlust von Informationen. Und mit dem Verlust von Information vermindert

sich auch zwangsläufig die Transparenz der gefundenen Ergebnisse. Man muss sich Folgendes vor Augen führen. Die durchschnittliche Bearbeitungszeit einer Studie für die ICCT dauerte circa 1,5 Stunden. Dabei alle relevanten Informationen zu den einzelnen Schlüsselpunkten zu finden erforderte große Geduld und war nicht immer einfach. Während die definierten Endpunkte meistens als solche gekennzeichnet waren, erforderte es ein sehr aufmerksames Lesen der Diskussion um für alle definierten Endpunkte die beobachteten Ergebnisse zu finden. Dabei wurden viele Ergebnisse oft nur in einem Nebensatz erwähnt. Dadurch war die Chance sehr groß, dass diese Ergebnisse einfach überlesen wurden.
Eine Originalarbeit besteht zudem aus mehreren Seiten Text mit einer Fülle an Informationen. Welcher normale Leser kann sich dabei sämtliche definierten Endpunkte merken? Dass es dabei möglicherweise kaum auffällt, dass für den primären Endpunkt einer Studie kein Ergebnis beobachtet wurde, ist leicht vorzustellen.

Durch die fehlende inhaltliche Konsistenz klinischer Studien ist es bisher nicht möglich eine bestimmte Stelle im Text herauszunehmen, die alle wichtigen Informationen enthält. Wer nur den Abstract einer Studie liest, dem entgehen unter Umständen wichtige Informationen. Das konnte in vielen Arbeiten gezeigt werden. [43][45] Der Befund der ICCT, dass sich die Schlussfolgerung im Abstract in fünf Studien nicht auf das Ergebnis des primären Endpunktes bezog, kann das nur unterstreichen.

Ko et al forderten 2002 die Berichterstattung klinischer Studien zu verbessern. [62] Eine der Konsequenzen war die Weiterentwicklung der CONSORT Kriterien. [4] Wenn es also keine Stelle in einer klinische Studie gibt, die alle relevanten Informationen der Studie enthält, dann zieht die ICCT die Konsequenz, dass eine solche Stelle geschaffen werden muss.

Jede Studie trifft eine Aussage über den Effekt einer Intervention an der Studiengruppe. Dazu definiert sie klinische Endpunkte. Eine Lösung zum Problem der inhaltlichen Konsistenz und dem damit einhergehenden Verlust an Informationen ist die Einführung einer neuen Tabelle. In dieser Tabelle sind alle Informationen aus der Studie enthalten. Sie beschreibt die Intervention, enthält alle Informationen zur Studiengruppe und alle definierten Endpunkte. Der Studiengruppe wird die Empfehlungsgruppe gegenübergestellt. Diese entspricht dem Patientenkollektiv, auf das die gefundenen Ergebnisse Anwendung finden. Zuletzt wird jedem definierten Endpunkt das entsprechend beobachtete Ergebnis gegenübergestellt.

Die vorgeschlagene Tabelle könnte folgende Form besitzen.

Intervention	
• *In diesem Abschnitt wird kurz die durchgeführte Intervention beschrieben. Es werden alle relevanten Informationen aufgezeigt*	
Study Population	***Population of Recommendation***
• *Es werden kurz die Einschluss/Ausschlusskriterien dargestellt*	• *Hier sollen die Charakteristika derjenigen Population stehen, für die eine Empfehlung bezüglich der Intervention ausgesprochen wird*
Predefined Outcomes	***Results***
Primary Outcome(s)	**First Result(s)**
• *Hier wird der primäre Endpunkt stehen*	• *Direkt daneben steht in Kurzform, ohne Wertung, ein gefundenes Ergebnis*
Secondary Outcome(s)	**Second Result(s)**
• *Gleiches gilt für den sekundären Endpunkt*	• *Ergebnis in Kurzform*
Additional Outcome(s)	**Additional Result(s)**
• *Gleiches gilt für zusätzliche Endpunkte*	• *Ergebnis in Kurzform*

Abbildung 16: Beispiel einer möglichen neuen Tabelle mit allen notwendigen Informationen

Eine Tabelle bietet einige Vorteile. Die durchgeführte Intervention erschließt sich auf einen Blick. Das Gleiche gilt für die Studienpopulation und die ihr gegenübergestellte Empfehlungsgruppe. Der entscheidende Vorteil besteht jedoch in der Gegenüberstellung der Endpunkte und der beobachteten Ergebnisse. Das mühsame Suchen der Endpunkte entfällt, ebenso das manchmal kaum mögliche Auffinden der beobachteten Ergebnisse. Liegen alle Endpunkte und deren Ergebnisse auf einen Blick vor, dann können keine Informationen mehr verloren gehen. Die Editoren klinischer Journals könnten durch den Vergleich des Studienprotokolls mit dieser Tabelle leicht prüfen, ob die Originalarbeit dem zugrunde liegenden Studienprotokoll entspricht. Dann wäre es einfacher, mögliche Fehlerquellen zu finden.

Klinische Studien sind das Herzstück der Evidenzbasierten Medizin. [2] Ihre Ergebnisse sind zur entscheidenden Grundlage für das tägliche ärztliche Handeln geworden. Sollte es also nicht eine Selbstverständlichkeit sein, dass man diesen Ergebnissen auch trauen kann? Man denke an einige der großen Arzneimittelskandale der letzten Jahre. Medikamente zur Senkung des Cholesterins wurden zu Millionen in der Patientenversorgung eingesetzt, um dadurch die Folgen der Arteriosklerose zu mildern. Studien wiesen dabei nach, dass der Nutzen größer sei als die Schädigung durch Nebenwirkungen. Im Endeffekt musste man aber feststellen, dass das so nicht stimmte. Gleiches passierte bei Mitteln zur Behandlung rheumatischer

Schmerzen. Sie wurden millionenfach verordnet, da sie zum einen die Schmerzen effektiv bekämpften und zum anderen weniger Nebenwirkungen aufwiesen als alte Medikamente mit ähnlicher Wirkung. Auch hier wiesen klinische Studien eine gute Wirksamkeit bei geringem Nebenwirkungspotential nach. Dass diese Medikamente das Risiko eines Herzinfarktes erhöhten, bemerkte man erst später.

Gesundheit ist ein hohes Gut. Sie nach bestem Wissen und Gewissen zu wahren die Aufgabe jedes klinisch tätigen Arztes. Um seine Aufgabe zu erfüllen, muss der Arzt den Ergebnissen der medizinischen Forschung vertrauen können. Medizinische Forschung erfolgt nicht zum Selbstzweck. Sie steht im Dienste am Menschen und sie geschieht für den Menschen. Darum sollte sie nicht zum Spielball finanzieller Interessen werden. Dabei ist sie das schon längst geworden. [36][38][58][59]

Es gibt positive Ansätze, die das Vertrauen in die Ergebnisse klinischer Studien stärken. Man denke an das Ottawa Statement oder an die Einführung der CONSORT Kriterien. [4][39] Raum für Verbesserungen bleibt. Bereits 2007 forderten *Laine et al* eine medizinische Forschung, der die Öffentlichkeit wieder trauen kann. [42] Diesem Ziel müssen wir weiter folgen.

5. Zusammenfassung

„The Internal Consistency of Clinical Trials“, die ICCT, untersucht die inhaltliche Konsistenz randomisierter klinischer Studien. Setzt man voraus, dass klinische Studien in ihrem Inhalt konsistent sind, dann kann man die Schlussfolgerung ziehen, dass es möglich ist eine Stelle im Text einer Studie zu finden, die alle relevanten Informationen der Arbeit enthält. Zeigt die ICCT, dass klinische Studien in ihrem Inhalt nicht konsistent sind dann können zum einen die Ursachen dargestellt werden, zum anderen kann die ICCT Wege aufzeigen um die Berichterstattung randomisierter klinischer Studien zu verbessern. Um die inhaltliche Konsistenz klinischer Studien zu überprüfen, sammelt die ICCT alle verfügbaren Informationen zu definierten Schlüsselpunkten an verschiedenen Textstellen der Originalstudien und vergleicht diese Informationen miteinander. Stimmen alle Informationen überein, dann sind die Schlüsselpunkte an den jeweiligen Textstellen inhaltlich konsistent. Als Textstellen vergleicht die ICCT den Titel, die Einleitung, die Methodik, die Diskussion und den Abstract. Insgesamt werden 53 Studien aus drei Journals für diese Arbeit untersucht. Die ICCT unterteilt jede klinische Studie in vier Ebenen. In der ersten Ebene wird die Einleitung der Studie mit der Methodik verglichen. Die zweite Ebene vergleicht zum einen die Methodik mit der Diskussion und zum anderen die Diskussion mit dem Abstract. In der dritten Ebene vergleicht die ICCT den Titel mit der Methodik und der Diskussion. Die vierte Ebene untersucht, ob sich die Schlussfolgerung im Abstract auf das Ergebnis für den primären Endpunkt bezieht. Die Datenauswertung folgt einem festen Schema und ist rein numerisch. Zuerst werden fünf Hypothesen geprüft. Diese postulieren, dass die Textstellen in allen Schlüsselpunkten der jeweiligen Ebenen übereinstimmen. Dann wird die Anzahl der Studien bestimmt, die in einer der ersten drei Ebenen nicht übereingestimmt haben. Als nächstes wird die Anzahl der Nichtübereinstimmungen in den einzelnen Schlüsselpunkten der ersten drei Ebenen gezählt. Schließlich werden die Gründe dafür aufgezählt, dass der Schlüsselpunkt *DAS ERGEBNIS* in den Textstellen Einleitung, Methodik, Diskussion und Abstract nicht übereingestimmt hat. Die Auswertung der vierten Ebene erfolgt zuletzt.

Als Ergebnis fand die ICCT heraus, dass keine der untersuchten 53 Studien in ihrem Inhalt konsistent waren. Die Einleitung stimmte in 26 Studien nicht mit der Methodik überein. In 36 Studien unterschied sich die Diskussion von der Methodik. Der Vergleich zwischen der Diskussion und dem Abstract ergab, dass 46 Studien nicht übereinstimmten. Die durch den Titel geweckten Erwartungen an die Intervention wurden in 49 Studien erfüllt. In 42 Studien stimmte die Fragestellung im Titel mit den beobachteten Ergebnissen aus der Diskussion überein. Trennte man diese Ergebnisse nach den einzelnen Schlüsselpunkten dann unterschied

sich die Studiengruppe in der Einleitung in drei Studien von jener der Methodik. Die Intervention stimmte an diesen beiden Textstellen in sechs Studien nicht überein. Die Ziele der Intervention aus der Einleitung und die definierten Endpunkten der Methodik unterschieden sich in 24 Studien. Verglich man die Studiengruppe in der Methodik und der Diskussion, unterschied sie sich in 15 Studien. Die definierten Endpunkte in der Methodik stimmten in 36 Studien nicht mit den beobachteten Ergebnissen der Diskussion überein. Bei neun Studien stimmten die Studiengruppen im Abstract und der Diskussion nicht überein. Die beobachteten Ergebnisse der Diskussion unterschieden sich in 46 Studien von den Ergebnissen im Abstract. Als Gründe für die fehlende inhaltliche Übereinstimmung des Schlüsselpunktes *DAS ERGEBNIS* fanden sich in der ersten Ebene, dass sieben Studien in der Einleitung eine Fragestellung formulierten, ohne dafür einen Endpunkt zu definieren. 16 Studien definierten in der Methodik einen Endpunkt ohne eine entsprechende Fragestellung in der Einleitung. In der zweiten Ebene wurden bei 21 Studien in der Diskussion Ergebnisse beobachtet für die in der Methodik kein Endpunkt definiert worden war. Sieben Studien definierten einen primären Endpunkt, beobachteten dafür jedoch kein Ergebnis. In 12 Studien wurde für sekundäre Endpunkte kein Ergebnis beobachtet. 15 Studien definierten zusätzliche Endpunkte, ohne dafür ein Ergebnis zu beobachten. Im Vergleich zwischen Abstract und der Diskussion fand sich für 15 Studien ein Ergebnis im Abstract ohne entsprechendes Ergebnis in der Diskussion. In zwei Studien fehlte das beobachtete Ergebnis für den primären Endpunkt im Abstract, in 17 Studien fehlten beobachtete Ergebnis für sekundäre und in 18 Studien für zusätzliche Endpunkte. Zuletzt fand sich in 13 Studien ein beobachtetes Ergebnis nur in der Diskussion und fehlte sowohl in der Methodik als auch im Abstract. In der vierten Ebene fand man heraus, dass sich in 31 Studien die Schlussfolgerung im Abstract direkt auf das Ergebnis des primären Endpunktes bezog.

Die ICCT konnte zeigen, dass randomisierte klinische Studien in ihrem Inhalt nicht konsistent sind. Daraus kann man die Schlussfolgerung ziehen, dass es in einer klinischen Studie keine Textstelle gibt, die alle relevanten Informationen zu der jeweiligen Studie enthält. Um alle Informationen zu erhalten, muss die gesamte Originalarbeit gelesen werden. Die Präsentation der Ergebnisse von Studien wurde als eigentliches Problemfeld identifiziert. Der größte Informationsverlust geschieht an diesem Schlüsselpunkt. Als Konsequenz daraus muss die Berichterstattung klinischer Studien verbessert werden, damit ein Verlust von relevanten Informationen vermieden wird. Dies fördert die Transparenz der gefundenen Ergebnisse und macht es somit einfacher den Ergebnissen klinischer Studien zu trauen.

6. Literaturverzeichnis

1. W. Eckart: Geschichte der Medizin *Springer Verlag, 6. Auflage 2008*
2. B. Sibbald, M. Roland: Understanding controlled trials: Why are randomised controlled trials important? *Br Med J 1998;316: 201*
3. D. L. Sackett**,** W. Rosenberg**,** J.A. Muir Gray**,** R.B. Haynes**,** W. Richardson: Evidence based medicine: What it is and what it isn't. *Br Med J 1996;312:71-72*
4. D. Moher, K. Schulz, D. Altman: The CONSORT statement: Revised recommendations for improving the quality of reports of parallel group randomized trials. *Lancet 2001;357:1191-1194*
5. D.L Sackett, W.S. Richardson, W. Rosenberg, R.B. Haynes: Evidenzbasierte Medizin. *W. Zuckschwerdt Verlag, München*
6. D. Coultas: Ethical Considerations in the Interpretation and Communication of Clinical Trial Results. *Proc Am Thorac Soc 2007;4: 194-199*
7. K.B. Freeman, S. Back, J. Bernstein: Sample Size and statistical power of randomised, controlled trials in orthopaedics. *J Bone Joint Surg Br 2001;83:397-402*
8. D. McNamee, A. James, S. Kleinert: Protocol Review at The Lancet. *Lancet 2008;372:189-190*
9. J.C. Sánchez-Thorin, M.C. Córtes, M. Montenegro, N. Villate: The quality of reporting of randomized clinical trials published in Ophthalmology. *Ophthalmology 2001;108:410-415*
10. J.B. Hrachovec, M. Mora, J.M. Wright et al.: Reporting of 6-Month vs. 12-Month Data in an Clinical Trial of Celecoxib. *J Am Med Assoc 2001;286:2398-2400*
11. S. Hahn, P.R. Williamson, J.L. Hutton: Investigation of within-study selective reporting in clinical research: follow-up of applications submitted to a local research ethics committee. *J Eval Clin Pract 2002;8:353-359*
12. A.W. Chan, A. Hróbjartsson, M.T. Haahr, P.C. Gotzsche, D.G. Altman: Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *J Am Med Assoc 2004;291:2457-2465*
13. K. Chalkidou, J. Lord, A. Fischer, P. Littlejohns: Evidence-Based Decision Making: Should we wait for more information? *Health Aff 2008;27:1642-1653*
14. A.J. Vickers: Underpowering in randomized trials reporting a sample size calculation. *J Clin Epidemiol 2003;56:717-720*
15. D. Moher, C.S. Dulberg, G.A. Wells: Statistical power, sample size, and their reporting in randomized controlled trials *J Am Med Assoc 1994;27:122-124*

16. K. Dickersin, D. Rennie: Registering Clinical trials. *J Am Med Assoc 2003:290:516-523*
17. K.B. Chan, M. Man-Son-Hing, F.J. Molnar, A. Laupacis: How well is the clinical importance of study results reported? An assessment of randomized controlled trials. *C Med Assoc J 2001;165:1197-1202*
18. B. Als-Nielsen, W. Chen, C. Gluud, L. L. Kjaergard: Association of Funding and Conclusions in Randomized Drug Trials: A Reflection of Treatment Effect or Adverse Events? *J Am Med Assoc. 2003;290:921-928*
19. L.L. Gluud, T.I. Sorensen, P.C. Gotzsche, C. Gluud: The journal impact factor as a predictor of trial quality an outcomes: cohort study of hepatobiliary randomized clinical trials. *Am J Gastroenterol 2005;100:2431-2435*
20. N.A. Desbiens: The presence of hypotheses in the medical literature. *Am J Med Sci 2004;328:319-322*
21. J.M. Lachin: Statistical considerations in the intent-to-treat principle. *Control Clin Trials 2000;21:167-189*
22. G. Jones, K. Abbasi: Trial protocols at the BMJ. *Br Med J 2004;329:1360*
23. J. Pildal, A.W. Chan, A. Hróbjartsson, E. Forfang, D.G. Altman, P.C. Gotzsche: Comparison of description of allocation concealment in trial protocols and the published reports. *Br Med J 2005;330:1049*
24. S. Eldrige, C. Cryer, G. Feder, M. Underwood: Sample size calculations for intervention trials in primary care randomizing by primary care group: an empirical illustration from one proposed intervention trial. *Stat Med 2001;20:267-376*
25. K.F. Schulz, D.A. Grimes: Sample size calculations in randomised trials: mandatory and mystical. *Lancet 2005;365:1348-1353*
26. D. Fergussion, S.D. Aaron, G. Guyatt, P. Hébert: Post-randomisation exclusions: the intention to treat principle an excluding patients from analysis. *Br Med J 2002;325: 652-654*
27. S. Eldrigde, D. Ashby, C. Bennett, M. Wakelin, G. Feder: Internal and external validity of cluster randomised trials: systematic review of recent trials. *Br Med J 2008;336:876-880*
28. C.S. Weaver, J. Leonardi-Be, F. Bath-Hextall, P. Bath: Sample size calculation in Acute Stroke Trials: a systematic review of their reporting, characteristics and relationships with outcome. *Stroke 2004;35:1216-1224*

29. N. Porta, C. Bonet, E. Cobo: Discordance between reported intention-to-treat and per protocol analyses. *J Clin Epidemiol 2007;60:663-669*
30. R. Schiffner, J. Schiffner-Rohe, M. Gerstenhauer, F. Hofstädter, M. Landthaler, W. Stolz: Differences in efficacy between intention-to-treat and per-protocol analyses for patients with psoriasis vulgaris and atopic dermatitis: clinical and pharmacoeconomic implications. *Br J Dermatol 2002;144:1154-1160*
31. J. Gravel, L. Opatrny, S. Shapiro: The intention-to-treat approach in randomized controlled trials: are authors saying what they do and doing what they say? *Clin Trials 2007;4:350-356*
32. S. Hollis, F. Campbell: What is meant by intention to treat analysis? Survey of published randomised controlled trials. *Br Med J 1999;319:670-674*
33. D.G. Altman, K.F. Schulz, D. Moher, M. Egger, F. Davidoff, D. Elbourne, P.C. Goetzsche, T. Lang: The Revisec CONSORT Statement for reporting randomized trials: Explanation and Elaboration. *Ann Intern Med 2001;134:663-694*
34. H.P. Soares, S. Daniels, A. Kumar, M. Clarke, C. Scott, S. Swann, B. Djulbegovic: Bad reporting does not mean bad methods for randomised trials: observational study of randomised controlled trials performed by the radiation therapy oncology group. *Br Med J 2004;328:22-24*
35. C. DeAngelis: The influence of money on medical science. *J Am Med Assoc 2006:296:996-998*
36. P.C. Gotzsche, A. Hróbjartsson, H. Johansen et al.: Constraints on publication rights in industry-initiated clinical trials. *J Am Med Assoc 2006;295:1645-1646*
37. S. Lagakos: The challenge of Subgroup Analyses – Reporting without distorting. *N Eng J Med 354;16:1667-1670*
38. E. Turner, A. Matthews, E. Linardatos, R. Tell, R. Rosenthal: Selective Publication of Antidepressant Trials and its influence on apparent Efficacy. *N Eng J Med 2008;358: 252-260*
39. K. Krleza-Jeric, A-W. Chan, K. Dickersin, I. Sim, J. Grimshaw, C. Gluud: Principles for international registration of protocol information and results from human trials of health related interventions: Ottawa statement (part 1). *Br Med J 2005;330:956-958*
40. F.B. Karassa, A. Tatsioni, J.P. Ioannidis: Design, quality and bias in randomized controlled trials of systemic lupus erythematosus. *J Rheumatol 2003;30:979-984*
41. V. Montori, P.J. Devereaux, N. Adhikari et al.: Randomized trials stopped early for benefit: A systematic review. *J Am Med Assoc 2005;294:2203-2209*

42. C. Laine, S. Goodman, M. Griswold, H. Sox: Reproducible Research: moving toward research the public can really trust. *Ann Intern Med 2007;146:450-453*
43. M. Toma, F.A. McAlister, L. Bialy, D. Adams, B. Vandermeer, P.W. Armstrong: Transition from meeting abstract to full-length journal article for randomized controlled trials. *J Am Med Assoc 2006;295:1281-1287*
44. Y. Dundar, S. Dodd, R. Dickson, T. Walley, A. Haycox, P.R. Williamson: Comparison of conference abstracts and presentations with full-text articles in the health technology assessments of rapidly evolving technologies. *Health Technol Assess 2006;10:1-145*
45. R.W. Scherer, P. Langenberg, E. von Elm: Full Publication of results initially presented in abstracts. *Cochrane Database Syst Rev 2007;(2):MR000005*
46. R.W. Scherer, K. Dickersin, P. Langenberg: Full publication of results initially presented in abstracts. A meta-analysis. *J Am Med Assoc* 1994;272:158-162
47. S. Hopewell, M. Clarke, L. Askie: Reporting of trials presented in conference abstracts needs to be improved. *J Clin Epidemiol 2006;59:681-684*
48. A.V. Hernández, E. Boersma, G.D. Murray, J.D. Habbema, E.W. Steyerberg: Subgroup analyses in therapeutic cardiovascular clinical trials: are most of them misleading? *Am Heart J 2006;151:257-264*
49. A.V. Hernández, E.W. Steyerberg, G.S. Taylor, A. Marmarou, J.D. Habbema, A.I. Maas: Subgroup analysis and covariate adjustment in randomized clinical trials of traumatic brain injury: a systematic review. *Neurosurgery 2005;57:1244-1253*
50. S.F. Assmann, S.J. Pocock, L.E. Enos, L.E. Kasten: Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet 2000, 355:1064-1069*
51. S.T. Brookes, E. Whitley, T.J. Peters, P.A. Mulheran, M. Egger, G. Davey Smith: Subgroup analysis in randomised controlled trials: quantifying the risk of false-positives and false-negatives. *Health Technol Assess 2001;5:1-56*
52. A.W. Chan, D.G. Altman: Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. *Br Med J 2005;330:753*
53. A. Le Henanff, B. Giraudeau, G. Baron, P. Rayaud: Quality of reporting of noninferiority and equivalence randomized trials. *J Am Med Assoc 2006;295:1147-1151*
54. A-W. Chan, A. Hróbjartsson, K. Jorgensen, P. Gotzsche, D.G. Altman: Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols. *Br Med J 2008;337:1523-1526*

55. A. Hróbjartsson, A-W. Chan, M.T. Haahr, P.C. Gotzsche, D.G. Altman: Selective reporting of positive outcomes in randomised trials. A comparison of protocols with published reports. *Ugeskr Laeger 2005;167:3189-3191*
56. A.M. Wood, I.R. White, S.G. Thompson: Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clin Trials 2004;1:368-376*
57. E. Hemminki: Study of information submitted by drug companies to licensing authorities. *Br Med J 1980;280:833-836*
58. H. Melander, J. Ahlqvist-Rastad, G. Meijer, B. Beermann: Evidence b(i)ased medicin-selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *Br Med J 2003;326:1171-1173*
59. K. Lee, P. Bacchetti, I. Sim: Publications of clinical trials supporting successful new drug applications: a literature analysis. *PLos Med 2008:5;401-404*
60. M.K. Krzyzanowska, M. Pintilie, I.F. Tannock: Factors associated with failure to publish large randomized trials presented at an oncology meeting. *J Am Med Assoc 2003;290:495-501*
61. S. Greenland: Commentary: On “Quality in epidemiological research: should we be submitting papers before we have the results and submitting more hypothesis generating research?”. *Int J Epidemiol 2007;36:944-945*
62. C.Y. Ko, J. Sack, J.T. Chang, A, Fink: Reporting randomized controlled trials: where quality of reporting may be improved. *Dis Colon Rectum 2002;45:443-447*
63. M.J. Marsh, B.S. Hawkins: Publications from multicentre clinical trials: statistical techniques and accessibility to the reader. *Stat Med 1994;13:2393-2406*
64. K. Adetugbo, H. Williams: How well are randomized controlled trials reported in the dermatology literature? *Arch Dermatol 2000;136:381-385*
65. A-W. Chan, R. Upshur, J. Singh, D. Ghersi, F. Chapuis, D.G. Altman: Waiving confidentially for the greater good. *Br Med J 2006;332.1086-1089*
66. C. Laine, R. Horton, C. DeAngelis, J. Drazen, F. Frizelle, F. Godlee, C. Haug, P. Hébert, S. Kotzin, A. Marusic, P. Sahni: Clinical Trial Registration – Looking Back and Moving Ahead. *N Engl J Med 2007;356:2734-2736*
67. C. Begg, M. Cho, S. Eastwood, R. Horton, D. Moher, I. Olkin et al: Improving the quality of reporting of randomized controlled trials: The CONSORT statement. *J Am Med Assoc 1996;276:1339-1341*

68. A. Fill: Das Prinzip Spannung: Sprachwissenschaftliche Betrachtung zu einem universalen Phänomen. *Gunter Narr Verlag Tübingen 2003;77*
69. American Journal of Obstetrics and Gynecology: *http://journals.elsevierhealth.com/periodicals/ymob/home* (Stand 11.09.2009)
70. The Lancet: *http://www.thelancet.com* (Stand 11.09.2009)
71. New England Journal of Medicine: *http://wwwm.nejm.com* (Stand 11.09.2009)
72. Journal Citation Report - 2003 Science Edition: *www.bio21.bas.bg/ibf/IF2003.pdf* (Stand 11.09.2009)
73. N. Jachertz: 100. Deutscher Ärztetag: Orientierung in Zeiten des Umbruchs. *Dtsch Arztebl 1997; 94:1613-1614*
74. The Cardiac Arrhythmia Suppression Trial II Investigators: Effect of the antiarrhythmic agent moricizine on survival after myocardial infarction. *N Engl J Med 1992: 327: 227-233*
75. Bortz: Statistik für Human- und Sozialwissenschaftler. *Springer Verlag 6. Auflage 2005*
76. Thierolf: Kosten und Finanzierung pharmazeutischer Forschung und Entwicklung. *Springer Verlag 2. Auflage 2008*
77. Gesetz über den Verkehr mit Arzneimitteln (Arzneimittelgesetz - AMG) § 22 Absatz 1
78. Sozialgesetzbuch (SGB) Fünftes Buch (V) - Gesetzliche Krankenversicherung - (Artikel 1 des Gesetzes v. 20. Dezember 1988, BGBl. I S. 2477) §35b
79. .M. Greenhalgh, C. Brown, D. Epstein: Endovascular aneurysm repair versus open repair in patients with abdominal aortic aneurysm (EVAR trial 1): randomized controlled trial *Lancet 2005; 365: 2179–86*
80. J. E. Wagner, J. S. Thompson, S. L. Carter, N. A Kernan et. al: Effect of graft-versus-host disease prophylaxis on 3-yeardisease-free survival in recipients of unrelated donor bone marrow (T-cell Depletion Trial): a multi-centre, randomized phase II–III trial *Lancet 2005; 366: 733–41*
81. R. A. Patchell, P. A. Tibbs, W. F. Regine, R. Payne et. al: Direct decompressive surgical resection in the treatment of spinal cord compression caused by metastatic cancer: a randomised trial *Lancet 2005; 366: 643–48*
82. S. Zhang, S. Wang, Q. Li, S. Yao et al.: Capillary leak syndrome in children with C4A-deficiency undergoing cardiac surgery with cardiopulmonary bypass: a double-blind, randomised controlled study *Lancet 2005; 366: 556–62*

83. K. F. Rabe, E. D. Bateman, D. O'Donnell, S. Witte et al.: Roflumilast—an oral anti-inflammatory treatment for chronic obstructive pulmonary disease: a randomized controlled trial *Lancet 2005; 366: 563–71*

84. M. Bolla, H. van Poppel, L. Collette, P. van Cangh et al.: Postoperative radiotherapy after radical prostatectomy: a randomised controlled trial (EORTC trial 22911) *Lancet 2005; 366: 572–78*

85. S. D. Aaron, K. L Vandemheen, W.Ferris, D. Fergusson et al: Combination antibiotic susceptibility testing to treat exacerbations of cystic fibrosis associated with multiresistant bacteria: a randomised, double-blind, controlled clinical trial *Lancet 2005; 366: 463–71*

86. S. Harvey, D. A. Harrison, M. Singer, J. Ashcroft et al. : Assessment of the clinical effectiveness of pulmonary artery catheters in management of patients in intensive care (PAC-Man): a randomised controlled trial *Lancet 2005; 366: 472–77*

87. R. T. D. Oliver, M. D. Mason, G. M. Mead, H. von der Maase et al. : Radiotherapy versus single-dose carboplatin in adjuvant treatment of stage I seminoma: a randomised trial *Lancet 2005; 366: 293–300*

88. J. McIntyre, S. Robertson, E. Norris, R. Appleton et al. : Safety and efficacy of buccal midazolam versus rectal diazepam for emergency treatment of seizures in children: a randomised controlled trial *Lancet 2005; 366: 205–10*

89. M. W. Krucoff, S. W. Crater, D. Gallup, J. C. Blankenship et al. : Music, imagery, touch, and prayer as adjuncts to interventional cardiac care: the Monitoring and Actualisation of Noetic Trainings (MANTRA) II randomised study *Lancet 2005; 366: 211–17*

90. C. Witt, B. Brinkhaus, S. Jena, K. Linde et al. : Acupuncture in patients with osteoarthritis of the knee: a randomised trial *Lancet 2005; 366: 136–43*

91. P. W. Rose, A. Harnden, A. B. Brueggemann, R. Perera et al. : Chloramphenicol treatment for acute infective conjunctivitis in children in primary care: a randomised double-blind placebo-controlled trial *Lancet 2005; 366: 37–43*

92. T. J. Schnitzer, G. R. Burmester, E. Mysler, M. C. Hochberg et al. : Comparison of lumiracoxib with naproxen and ibuprofen in the Therapeutic Arthritis Research and Gastrointestinal Event Trial (TARGET), reduction in ulcer complications: randomised controlled trial *Lancet 2004; 364: 665–74*

93. H.M. Colhoun, D. J. Betteridge, P. N. Durrington, G. A. Hitman et al.: Primary prevention of cardiovascular disease with atorvastatin in type 2 diabetes in the

Collaborative Atorvastatin Diabetes Study (CARDS): multicentre randomised placebo-controlled trial *Lancet 2004; 364: 685–96*

94. N. E. Vain, E. G. Szyld, L. M. Prudent, T. E. Wiswell et al. : Oropharyngeal and nasopharyngeal suctioning of meconium-stained neonates before delivery of their shoulders: multicentre, randomised controlled trial *Lancet 2004; 364: 597–602*
95. G. Remuzzi, M. Lesti, E. Gotti, M. Ganeva et al. : Mycophenolate mofetil versus azathioprine for prevention of acute rejection in renal transplantation (MYSS): a randomised trial *Lancet 2004; 364: 503–12*
96. I. Gilron, J. M. Bailey, M.Ed. Dongsheng, R. R. Holden et al.: Morphine, Gabapentin, or Their Combination for Neuropathic Pain *N Engl J Med 2005;352:1324-34*
97. R. S. Bresalier, R. S. Sandler, H. Quan, J. A. Bolognese et al.: Cardiovascular Events Associated with Rofecoxib in a Colorectal Adenoma Chemoprevention Trial *N Engl J Med 2005;352:1092-102*
98. W. J. Morgan, E. F. Crain, R. S. Gruchalla, G. T. O'Connor et al. : Results of a Home-Based Environmental Intervention among Urban Children with Asthma *N Engl J Med 2004;351:1068-80*
99. C. C. Apfel, K. Korttila, M. Abdalla, H. Kerger et al. : A Factorial Trial of Six Interventions for the Prevention of Postoperative Nausea and Vomiting *N Engl J Med 2004;350:2441-51*
100. M. Prinssen, E. L.G. Verhoeven, J. Buth, P. W.M. Cuypers et al. : A Randomized Trial Comparing Conventional and Endovascular Repair of Abdominal Aortic Aneurysms *N Engl J Med 2004;351:1607-18*
101. J. H. Stone et al. *:* Etanercept plus Standard Therapy for Wegener's Granulomatosis *N Engl J Med 2005;352:351-61*
102. B. G. Feagan, G. R. Greenberg, G. Wild, R. N. Fedorak et al. : Treatment of Ulcerative Colitis with a Humanized Antibody to the *a*4*b*7 Integrin *N Engl J Med 2005;352:2499-507*
103. C. P. Cannon, E. Braunwald, C. H. McCabe, J. T. Grayston et al.: Antibiotic Treatment of Chlamydia pneumonia after Acute Coronary Syndrome *N Engl J Med 2005;352:1646-54*
104. G. E. Thwaites, N. D. Bang, N. H. Dung, H. T. Quy et al. : Dexamethasone for the Treatment of Tuberculous Meningitis in Adolescents and Adults *N Engl J Med 2004;351:1741-51*

105. M.N. Oxman, M.J. Levin, G.R. Johnson, K.E. Schmader et al. : A Vaccine to Prevent Herpes Zoster and Postherpetic Neuralgia in Older adults *N Engl J Med 2005;352:2271-84*

106. J. R. Korzenik, B. K. Dieckgraefe, J. F. Valentine, D. F. Hausman et al. : Sargramostim for Active Crohn's Disease *N Engl J Med 2005;352:2193-201*

107. E. S. Gragoudas, A. P. Adamis, E. T. Cunningham, M. Feinsod et al. : Pegaptanib for Neovascular Age-Related Macular Degeneration *N Engl J Med 2004;351:2805-16*

108. D. Oakes, I. Shoulson, K. Kieburtz, A. Rudolph et al. : Levodopa and the Progression of Parkinson's Disease *N Engl J Med 2004;351:2498-508*

109. N.D. Desai, E. A. Cohen, C. D. Naylor, S. E. Fremes : A Randomized Comparison of Radial-Artery and Saphenous-Vein Coronary Bypass Grafts N Engl J Med 2004;351:2302-9

110. S. S. Mou, B. P. Giroir, E. A. Molitor-Kirsch, S.R. Leonard et al. : Fresh Whole Blood versus Reconstituted Blood for Pump Priming in Heart Surgery in Infants *N Engl J Med 2004;351:1635-44*

111. W.W. Fawzi, G. I. Msamanga, D. Spiegelman, R. Wie et al. : A Randomized Trial of Multivitamin Supplements and HIV Disease Progression and Mortality *N Engl J Med 2004;351:23-32*

112. L. Gustafsson, I. Leijonhufvud, A. Aronsson, A-K. Mossberg et al. : Treatment of Skin Papillomas with Topical *a*-Lactalbumin–Oleic Acid *N Engl J Med 2004;350:2663-72*

113. S. D. Solomon, J. J.V. McMurray, M. A. Pfeffer, J. Wittes et al. : Cardiovascular Risk Associated with Celecoxib in a Clinical Trial for colorectal Adenoma Prevention *N Engl J Med 2005;352:1071-80*

114. N. A. Maskell, C. W.H. Davies, A. J. Nunn, E. Hedley et al. : U.K. Controlled Trial of Intrapleural Streptokinase for Pleural Infection *N Engl J Med 2005;352:865-74*

115. B. N. Singh, S. N. Singh, D. J. Reda, X. C.Tang et al. : Amiodarone versus Sotalol for Atrial Fibrillation *N Engl J Med 2005;352:1861-72*

116. E. Braunwald, M. J. Domanski, S. E. Fowler, N. L.Geller et al. : Angiotensin-Converting–Enzyme Inhibition in Stable Coronary Artery Disease *N Engl J Med 2004;351:2058-68*

117. H. Nelson, D. J. Sargent, H. S. Wieand, J. Fleshman et al. : A Comparison of Laparoscopically Assisted and Open Colectomy for Colon Cancer *N Engl J Med 2004;350:2050-9*

118. M. Emre, D. Aarsland, A. Albanese, E. J. Byrne et al. : Rivastigmine for Dementia Associated with Parkinson's Disease *N Engl J Med 2004;351:2509-18*

119. P. Ruggenenti, A. Fassi, A. Parvanova Ilieva, S. Bruno et al. : Preventing Microalbuminuria in Type 2 Diabetes *N Engl J Med 2004;351:1941-51*

120. J. C.W. Edwards, L. Szczepa´nski, J. Szechi´nski, A. Filipowicz-Sosnowska et al. : Efficacy of B-Cell–Targeted Therapy with Rituximab in Patients with Rheumatoid Arthritis *N Engl J Med 2004;350:2572-81*

121. R. E. Hershberger, R. C. Starling, H. J. Eisen, C-H. Bergh et al. : Daclizumab to Prevent Rejection after Cardiac Transplantation *N Engl J Med 2005;352:2705-13*

122. F. J. Torriani, M. Rodriguez-Torres,.J. K. Rockstroh, E. Lissen et al. : Peginterferon Alfa-2a plus Ribavirin for Chronic Hepatitis C Virus Infection in HIV-Infected Patients *N Engl J Med 2004;351:438-50*

123. J. M. Aldrighi, D. C. Quail, J. Levy-Frebault, F. Aguas et al. : Predictors of hot flushes in postmenopausal women who receive raloxifene therapy *Am J Obstet Gynekol 2004;191:1979-88*

124. M. Kühnert, S. Schmidt: Intrapartum management of nonreassuring fetal heart rate patterns: A randomized controlled trial of fetal pulse oximetry *Am J Obstet Gynekol 2004; 191:1989-95*

125. S. W. Leung, M. W. Pang, T. K. H. Chung: Retained products of gestation in miscarriage: An evaluation of transvaginal ultrasound criteria for diagnosing an ‘‘empty uterus’’ *Am J Obstet Gynekol 2004; 191:1133-7*

126. G. B. Bikhazi, M. C. Snabes, Z. H. Bajwa, D. J. Davis et al.: A clinical trial demonstrates the analgesic activity of intravenous parecoxib sodium compared with ketorolac or morphine after gynecologic surgery with laparotomy *Am J Obstet Gynekol 2004; 191:1183-91*

127. C. G. Sosa, E. Balaguer, J. G. Alonso, R. Panizza et al.: Meperidine for dystocia during the first stage of labor: A randomized controlled trial *Am J Obstet Gynekol 2004; 191:1212-8*

128. S. Gates, P. Brocklehurst, S. Ayers, U. Bowler et al.: Thromboprophylaxis and pregnancy: Two randomized controlled pilot trials that used low-molecularweight heparin *Am J Obstet Gynekol 2004; 191:1296-303*

129. A. Bisits, G. Madsen, M. Knox, A. Gill et al. : The Randomized Nitric Oxide Tocolysis Trial (RNOTT)for the treatment of preterm labor *Am J Obstet Gynekol 2004; 191:683-90*
130. A.J. Wolfberg, K. B. Michels, W. Shields, P. O'Campo et al. : Dads as breastfeeding advocates: Results from a randomized controlled trial of an educational Intervention *Am J Obstet Gynekol 2004; 191:708-12*
131. R. Williams, J. B. Hillard, G. De Vane, Tim Yeko et al. : A randomized, multicenter study comparing the efficacy of recombinant FSH vs recombinant FSH with Ganirelix during superovulation/IUI therapy *Am J Obstet Gynekol 2004; 191:648-53*

7. Anhang

- **Die Definitionen zu den Fundstellen der Schlüsselpunkte in den Textstellen**
- **Die Definitionen für inhaltliche Übereinstimmungen/Nichtübereinstimmungen und möglichen Übereinstimmungen der Textstellen:**

Für die erste Ebene –Die Intervention- wird definiert:

Fundstelle der Schlüsselpunkte im Text

Studiengruppe

Für die Einleitung gilt:

Am Ende der Einleitung ist diejenige Population als Studiengruppe zu werten, für die eine Intervention geplant ist, die mit einer bestimmten Fragestellung verknüpft wird.

Für die Methodik gilt:

Die tatsächliche, an dieser Studie teilnehmende Studiengruppe ist diejenige, welche in der Methodik als Studienpopulation definiert wird.

Die Intervention

Für die Einleitung gilt:

Die geplante Intervention innerhalb der Einleitung ist dasjenige medizinische Verfahren, welches mit einer bestimmten Fragestellung verknüpft in der vorliegenden Studie untersucht werden soll.

Für die Methodik gilt:

Die tatsächlich durchgeführte Intervention ist dasjenige medizinische Verfahren, welches, mit einer konkreten Fragestellung verknüpft, in der Methodik an der Studiengruppe durchgeführt wird.

Das Ziel der Intervention

Für die Einleitung gilt:

Das Ziel der Intervention ist diejenige Fragestellung am Ende der Einleitung die, bezogen auf eine geplante Intervention an einer geplanten Studienpopulation, beantwortet werden soll.

Für die Methodik gilt:

In der Methodik gelten all diejenigen Punkte als Endpunkte, die direkt als Endpunkte definiert werden, d.h. als primärer, sekundärer oder zusätzlicher Endpunkt. Ebenso gelten als zusätzliche Endpunkte und somit als weitere Fragestellung an diese Studie all diejenigen

Punkte oder Untersuchungen, die, mit einer Fragestellung verknüpft, eingeführt werden. Dies gilt auch dann, wenn die Studie diese Punkte nicht direkt als Endpunkte bezeichnet.

Was gilt als Übereinstimmung:

Schlüsselpunkt Studiengruppe:

- Die in der Einleitung und der Methodik als Studiengruppe identifizierten Schlüsselpunkte benennen namentlich und dem Sinn entsprechend die gleiche Gruppe. Wenn genannt müssen sowohl Alter und Geschlecht, als auch Komorbiditäten, Krankheitsbilder und andere spezifische Charakteristika identisch sein
- Eine Übereinstimmung wird auch dann angenommen, wenn die Studiengruppe aus der Einleitung in der Methodik genauer definiert wird, diese Definition der Studiengruppe aber keine grundsätzliche Änderung der Studiengruppe, wie in der Einleitung beschrieben, beinhaltet. Sollten also Hinweise auf Alter, Geschlecht, Krankheitsbilder, Komorbiditäten und andere spezifische Charakteristika der Studiengruppe in der Einleitung vorhanden sein, so dürfen sie in der Methodik genauer beschrieben werden, solange sich dadurch die grundsätzlichen Charakteristika der in der Einleitung beschriebenen Studiengruppe nicht ändern.

Schlüsselpunkt Intervention:

- Die in der Einleitung geplante Intervention entspricht der in der Methodik definierten. Weder die Einleitung noch die Methodik enthalten eine zusätzliche Intervention, die in der jeweils anderen Textstelle nicht genannt wird.
- Eine Übereinstimmung wird auch dann angenommen, wenn die Intervention in der Methodik genauer definiert wird, durch diese genauere Definition die durchgeführte Intervention in der Methodik jedoch im Grundsatz weiterhin jener geplanten Intervention aus der Einleitung entspricht. Im Grundsatz gleich bedeutet, dass zwar eine genauere Beschreibung von beispielsweise der Dauer der Anwendung, der Dosis und dem Zeitpunkt der Intervention in der Methodik erfolgen kann. Die Einführung neuer Substanzen und Verfahren oder die Aufsplitterung von Substanzen, die in der Einleitung als Kombinationstherapie eingeführten werden, in Einzelsubstanzen oder Einzelverfahren innerhalb der Methodik als neuem Interventionsarm ist hingegen nicht als Übereinstimmung zu werten.

Schlüsselpunkt Das Ziel der Intervention:

- Findet sich für jedes Ziel der Intervention in der Einleitung ein definierter Endpunkt im Methodikteil und umgekehrt, so gilt dieses als Übereinstimmung.
- Definiert die Methodik ein in der Einleitung definiertes Ziel der Intervention näher, so gilt dies ebenfalls als Übereinstimmung, solange sich dadurch aber nichts an der grundsätzlichen Fragestellung, wie aus der Einleitung erwartet, verändert.

Was gilt als mögliche Übereinstimmung:

Schlüsselpunkt Die Studiengruppe:

- Eine Studiengruppe wird in der Einleitung nicht explizit genannt, kann aber aus der Art des geplanten medizinischen Verfahrens geschlossen werden. Diese so geschlossene Studiengruppe stimmt grundsätzlich mit der in der Methodik definierten überein. Grundsätzlich bedeutet, dass die Basischarakteristika einer so geschlossenen Studiengruppe, wie Geschlecht, Krankheitsbilder, Komorbiditäten und andere spezifische Charakteristika, mit den Basischarakteristika der definierten Studiengruppe in der Methodik übereinstimmen müssen.
- Es kann nicht zweifelsfrei entschieden werden, ob die Charakteristika der Studiengruppe aus der Einleitung mit jenen der definierten Studiengruppe aus der Methodik übereinstimmen. Ebenso kann eine Nichtübereinstimmung nicht zweifelsfrei ausgeschlossen werden.

Schlüsselpunkt Die Intervention:

- Die Intervention wurde in der Einleitung ungenau beschrieben, kann aber aus dem Zusammenhang heraus erschlossen werden. Diese so erschlossene Intervention stimmt im Grundsatz mit der in der Methodik tatsächlich durchgeführten Intervention überein.
- Die Intervention für die Verum-Medikation stimmt in der Einleitung und der Methodik überein, aus der Einleitung ist jedoch nicht ersichtlich, dass diese Verum-Medikation gegen ein Placebo verglichen werden soll.
- Es kann nicht zweifelsfrei entschieden werden ob die geplante Intervention in der Einleitung mit der tatsächlich durchgeführten Intervention in der Methodik übereinstimmt. Ebenso kann eine Nichtübereinstimmung nicht zweifelsfrei ausgeschlossen werden.

Schlüsselpunkt Das Ziel der Intervention:

- Definiert die Einleitung keine exakte Fragestellung an dies Studie, diese Fragestellung kann aber aus der Einleitung geschlossen werden, so gilt es dann als eine mögliche Übereinstimmung, wenn diese so geschlossene Fragestellung den definierten Endpunkten in der Methodik entspricht. Dies gilt auch dann, wenn mehrere Endpunkte in der Methodik definiert wurden als aus der Einleitung geschlossen werden konnten.
- Benennt die Einleitung als Fragestellung an die betreffende Studie sehr ungenau definierte Ziele, wie beispielsweise die Effektivität oder Sicherheit einer Intervention, so gilt dies als mögliche Übereinstimmung, solange die in der Methodik definierten Endpunkte mit dieser allgemeinen Fragestellung in Einklang gebracht werden können.
- Es gilt als mögliche Übereinstimmung wenn nicht zweifelsfrei entschieden werden kann, ob das Ziel der Intervention mit den definierten Endpunkten in der Methodik übereinstimmt. Ebenso kann eine Nichtübereinstimmung nicht sicher ausgeschlossen werden.

Was gilt als Nichtübereinstimmung:

Schlüsselpunkt Die Studiengruppe:

- Genannte Charakteristika einer Studiengruppe aus der Einleitung wie Alter, Geschlecht, Krankheitsbilder, Komorbiditäten oder andere spezifische Charakteristika finden sich in der Studiengruppe der Methodik nicht wieder oder widersprechen diesen gar. Ebenso gilt der umgekehrte Fall als Nichtübereinstimmung, wenn zusätzlich genannte Charakteristika der Studiengruppe der Methodik nicht mehr den grundsätzlichen Basischarakteristika der Studiengruppe der Einleitung entsprechen.
- Definiert die Einleitung eine Studiengruppe exakter als die Methodik, so gilt dies ebenfalls als eine Nichtübereinstimmung.

Schlüsselpunkt Die Intervention:

- Wird die Intervention aus der Einleitung durch eine zusätzliche Intervention in der Methodik ergänzt und umgekehrt, so gilt diese als Nichtübereinstimmung. Dies gilt jedoch nicht, wenn die zusätzliche Intervention in der Methodik allein in der Einführung eines Placebos besteht.
- Benennen die Einleitung und der Methodikteil unterschiedliche Interventionen, so gilt dies als Nichtübereinstimmung.

- Benennen die Einleitung und der Methodikteil die gleiche Intervention, enthält jedoch eine der beiden Textstellen eine Modifikation, welche den Basischarakter der Intervention verändert, so gilt dies ebenfalls als Nichtübereinstimmung. Unter einer Änderung des Basischarakters wird beispielsweise verstanden, dass als Kombination eingeführte Verfahren oder Therapien in Einzelsubstanzen bzw. Einzelverfahren auf gespalten werden.

Schlüsselpunkt Ziel der Intervention

- Benennt die Einleitung ein Ziel der Intervention für welches die Methodik keinen Endpunkt definiert, so gilt dies als Nichtübereinstimmung.
- Definiert die Methodik einen Endpunkt, dem kein Ziel der Intervention in der Einleitung entspricht, so gilt dies als Nichtübereinstimmung.

<u>Für die zweite Ebene –Das Ergebnis- Ebene E2A wird definiert:</u>

Die Fundstelle der Schlüsselpunkte im Text

Die Studiengruppe

<u>Für die Methodik gilt:</u>

Die tatsächliche, an dieser Studie teilnehmende Studienpopulation ist diejenige, welche in der Methodik als Studienpopulation definiert wird.

<u>Für die Diskussion gilt:</u>

Die „Empfehlungsgruppe“ ist diejenige Population für welche in der Diskussion, basierend auf der durchgeführten Intervention und der in der Studie beantworteten Fragestellungen, eine direkte, sich auf die gefundenen Ergebnisse stützende, Empfehlung gegeben wird.

Das Ergebnis

<u>Für die Methodik gilt:</u>

In der Methodik gelten all diejenigen Punkte als Endpunkte, die direkt als Endpunkte definiert werden, d.h. als primärer, sekundärer oder zusätzlicher Endpunkt. Ebenso gelten als zusätzliche Endpunkte und somit als weitere Fragestellung an diese Studie all diejenigen Punkte oder Untersuchungen, die, mit einer Fragestellung verknüpft, eingeführt werden. Dies gilt auch dann, wenn die Studie diese Punkte nicht direkt als Endpunkte bezeichnet.

Für die Diskussion gilt:

In der Diskussion werden alle Punkte als beobachtetes Ergebnis dieser Studie gewertet, welche zweifelsfrei erkennen lassen, dass es sich um ein in dieser Studie gefundenes Ergebnis handelt. Dies ist dann der Fall, wenn sich aus Satzbau, Satzsinn und Zusammenhang des Textes erkennen lässt, dass der betreffende Satz direkt von einem Ergebnis, oder einer Schlussfolgerung zu einem Ergebnis, handelt.

Was gilt als Übereinstimmung:

Schlüsselpunkt Die Studiengruppe:

- Als Übereinstimmung gilt, wenn die in der Methodik definierte Studiengruppe hinsichtlich ihrer Basischarakteristika der Studiengruppe in der Diskussion entspricht. Das bedeutet, dass sowohl Alter, Geschlecht, Krankheitsbilder, Komorbiditäten als auch andere spezifische Charakteristika, wenn genannt, identisch sein müssen.
- Die Studiengruppe in der Diskussion kann weniger genau definiert sein als dies in der Methodik der Fall ist, solange die Basischarakteristika beider Gruppen übereinstimmen und wichtige Ergänzungen, wie beispielsweise bestimmte Ausschlusskriterien für die Studiengruppe in der Methodik, auch für die Studiengruppe in der Diskussion erkennbar sind.

Schlüsselpunkt Das Ergebnis:

- Eine Übereinstimmung gilt dann als gegeben, wenn für jeden definierten Endpunkt in der Methodik ein beobachtetes Ergebnis in der Diskussion zu finden ist. Ebenso muss für jedes beobachtete Ergebnis in der Diskussion ein entsprechender Endpunkt in der Methodik definiert worden sein.
- Findet sich ein beobachtetes Ergebnis in der Diskussion , für welches in der Methodik kein expliziter Endpunkt definiert wurde, ein solcher Endpunkt kann aber aus verschiedenen vorgestellten Untersuchungen und Messmethoden in der Methodik sicher geschlossen werden, dann gilt dies ebenfalls als Übereinstimmung

Was gilt als mögliche Übereinstimmung:

Schlüsselpunkt Die Studiengruppe:

- Ist aus der Diskussion nicht eindeutig ersichtlich für welche Population Empfehlungen ausgesprochen oder Konsequenzen gezogen werden, diese Population kann jedoch aus dem Zusammenhang heraus geschlossen werden, so gilt dies als mögliche Übereinstimmung wenn diese so geschlossene Studiengruppe aus der Diskussion in ihren Basischarakteristika jener in der Methodik definierten Studiengruppe nicht widerspricht.
- Es kann nicht zweifelsfrei entschieden werden, ob die Charakteristika der Studiengruppe aus der Diskussion mit jenen der definierten Studiengruppe aus der Methodik übereinstimmen. Ebenso kann eine Nichtübereinstimmung nicht zweifelsfrei ausgeschlossen werden.

Schlüsselpunkt Ergebnis:

- Kann für ein beobachtetes Ergebnis in der Diskussion in der Methodik kein expliziter Endpunkt gefunden werden, ein solcher Endpunkt aber aus verschiedenen vorgestellten Untersuchungen und Messmethoden geschlossen werden, allerdings nicht zweifelsfrei, dann gilt dies als mögliche Übereinstimmung.
- Es gilt als mögliche Übereinstimmung wenn nicht zweifelsfrei festgestellt werden kann, ob für jeden definierten Endpunkt in der Methodik ein entsprechendes beobachtetes Ergebnis in der Diskussion vorhanden ist und umgekehrt, dieses aber ebenso nicht sicher ausgeschlossen werden kann.

Was gilt als Nichtübereinstimmung:

Schlüsselpunkt Die Studiengruppe:

- Unterscheiden sich die Studiengruppe aus der Methodik und der Diskussion in wichtigen Basischarakteristika wie Alter, Geschlecht, Krankheitsbildern, Komorbiditäten und anderen spezifischen Charakteristika, so gilt dies als Nichtübereinstimmung.
- Der Schlüsselpunkt Studiengruppe stimmt an den beiden Textstellen auch dann nicht überein, wenn zwar die Basischarakteristika der Studiengruppe in der Methodik und in der Diskussion identisch sind, wichtige Ausschlusskriterien für die

Studienpopulation in der Methodik aus der Empfehlungsgruppe in der Diskussion aber nicht mehr ersichtlich sind.

Schlüsselpunkt Das Ergebnis:

- Findet sich für einen definierten Endpunkt aus der Methodik kein beobachtetes Ergebnis in der Diskussion wieder, so gilt dieser Fall als Nichtübereinstimmung. Es gilt auch dann als Nichtübereinstimmung wenn in der Diskussion ein Ergebnis beobachtet wird, für das in der Methodik kein Endpunkt definiert ist.

<u>Für die zweite Ebene –Das Ergebnis- Ebene E2B wird definiert:</u>

Die Fundstelle der Schlüsselpunkte im Text

Die Studiengruppe

<u>Für die Diskussion gilt:</u>

Die „Empfehlungsgruppe“ ist diejenige Population, für welche in der Diskussion , basierend auf der durchgeführten Intervention und der in der Studie beantworteten Fragestellungen, eine direkte, sich auf die gefundenen Ergebnisse stützende, Empfehlung gegeben wird.

<u>Für den Abstract gilt:</u>

Im Abstract gilt jene Population als Studiengruppe, für welche sich in der Schlussfolgerung des Abstracts eine, auf die beobachteten Ergebnisse der Studie, basierende Empfehlung oder Konsequenz wieder findet. Nennt die Schlussfolgerung des Abstracts keine Studiengruppe, so gilt nachfolgend diejenige Population als Studiengruppe im Abstract, welche im Abschnitt Methodik des Abstracts als Studienpopulation definiert wurde.

Das Ergebnis

<u>Für die Diskussion gilt:</u>

In der Diskussion werden all jene Punkte als beobachtetes Ergebnis dieser Studie gewertet, welche zweifelsfrei erkennen lassen, dass es sich um ein in dieser Studie gefundenes Ergebnis handelt. Dies ist dann der Fall, wenn sich aus Satzbau, Satzsinn und Zusammenhang des Textes erkennen lässt, dass der betreffende Satz direkt von einem Ergebnis, oder einer Schlussfolgerung zu einem Ergebnis, handelt

Für den Abstract gilt:

Als beobachtetes Ergebnis im Abstract gelten alle Ergebnisse, die im Ergebnisteil des Abstracts genannt werden. Ebenso gilt als beobachtetes Ergebnis dasjenige Ergebnis, worauf sich die Schlussfolgerung im Abstract bezieht.

Was gilt als Übereinstimmung:

Schlüsselpunkt Die Studiengruppe:

- Als Übereinstimmung gilt wenn die im Abstract definierte Studiengruppe hinsichtlich ihrer Basischarakteristika jener Studiengruppe in der Diskussion entspricht. Das bedeutet, dass sowohl Alter, Geschlecht, Krankheitsbilder, Komorbiditäten und andere spezifische Charakteristika, wenn genannt, identisch sein müssen.
- Es gilt auch dann als Übereinstimmung, wenn die Studiengruppe im Abstract genauer definiert wird, sich die Basischarakteristika der Studiengruppe aus dem Abstract und der Diskussion jedoch nicht widersprechen. Es gilt ebenso der umgekehrt Fall, wenn die Studiengruppe in der Diskussion differenzierter beschrieben ist. Jedoch müssen an einer Textstelle explizit genannte Ausschlusskriterien und anderweitige Einschränkungen für die Studiengruppe auch an der jeweils anderen Textstelle beschrieben sein.

Schlüsselpunkt Das Ergebnis:

- Es gilt als Übereinstimmung, wenn für jedes beobachtete Ergebnis in der Diskussion ein entsprechendes Ergebnis im Abstract vorhanden ist. Ebenso muss für jedes Ergebnis aus dem Abstract ein entsprechendes beobachtetes Ergebnis in der Diskussion vorhanden sein.

Was gilt als mögliche Übereinstimmung:

Schlüsselpunkt Die Studiengruppe:

- Ist aus der Diskussion nicht eindeutig ersichtlich für welche Population Empfehlungen ausgesprochen oder Konsequenzen gezogen werden, diese Population kann jedoch aus dem Zusammenhang heraus geschlossen werden, so gilt dies als mögliche Übereinstimmung wenn diese so geschlossene Studiengruppe aus der Diskussion in

ihren Basischarakteristika jener im Abstract definierten Studiengruppe nicht widerspricht. Dies gilt ebenso für den umgekehrten Fall.

- Es kann nicht zweifelsfrei entschieden werden, ob die Charakteristika der Studiengruppe aus der Diskussion mit jenen der definierten Studiengruppe aus dem Abstract übereinstimmen. Ebenso kann eine Nichtübereinstimmung nicht zweifelsfrei ausgeschlossen werden.

Schlüsselpunk Das Ergebnis:

- Es gilt als mögliche Übereinstimmung wenn nicht zweifelsfrei festgestellt werden kann, ob für jedes beobachtete Ergebnis in der Diskussion ein entsprechendes Ergebnis im Abstract vorhanden ist und umgekehrt, dieses kann aber ebenso nicht sicher ausgeschlossen werden.

Was gilt als Nichtübereinstimmung:

Schlüsselpunkt Die Studiengruppe:

- Unterscheiden sich die Studiengruppe aus dem Abstract und der Diskussion in wichtigen Basischarakteristika wie Alter, Geschlecht, Krankheitsbildern, Komorbiditäten und anderen spezifischen Charakteristika, so gilt dies als Nichtübereinstimmung.
- Der Schlüsselpunkt Studiengruppe stimmt an den beiden Textstellen auch dann nicht überein, wenn zwar die Basischarakteristika der Studiengruppe im Abstract und in der Diskussion identisch sind, wichtige Ausschlusskriterien für die Studienpopulation im Abstract aus der Empfehlungsgruppe in der Diskussion aber nicht mehr ersichtlich sind und umgekehrt.

Schlüsselpunkt Das Ergebnis:

- Wurde für ein beobachtetes Ergebnis in der Diskussion kein entsprechendes Ergebnis im Abstract gefunden, so gilt dies als Nichtübereinstimmung.
- Gleiches gilt, wenn sich für Ergebnisse aus dem Abstract keine entsprechenden beobachteten Ergebnisse in der Diskussion wieder finden.
- Widersprechen sich Ergebnisse in ihrer Aussage im Abstract und in der Diskussion so gilt dieses ebenfalls als Nichtübereinstimmung.

Für die dritte Ebene –Die durch den Titel geweckten Erwartungen- wird definiert:

Die notwendigen Definitionen für die dritte Ebene werden zusammen vorgestellt und nicht getrennt für die Ebenen E3A und E3B beschrieben. Im Gegensatz zur zweiten Ebene ist eine getrennte Definition nicht notwendig, da die Ebenen E3A und E3B aus unterschiedlichen Schlüsselpunkten bestehen.

Die Fundstelle der Schlüsselpunkte im Text

Die Intervention

Für den Titel gilt:

Die geplante Intervention ist dasjenige medizinische Verfahren, dessen Anwendung durch den Titel der Studie angekündigt wird und somit in dieser Studie auf eine gezielte Fragestellung hin untersucht werden soll.

Für die Methodik gilt:

Die tatsächlich durchgeführte Intervention ist dasjenige medizinische Verfahren, welches in der Methodik als durchgeführt beschrieben worden ist.

Das Ergebnis

Für den Titel gilt:

Das aus dem Titel zu erwartende Ergebnis ist dasjenige Ergebnis, welches entweder als direkte Fragestellung bereits im Titel genannt wurde, oder aber aus dem Zusammenhang heraus sicher geschlossen werden kann.

Für die Diskussion gilt:

In der Diskussion werden all jene Punkte als beobachtetes Ergebnis dieser Studie gewertet, welche zweifelsfrei erkennen lassen, dass es sich um ein in dieser Studie gefundenes Ergebnis handelt. Dies ist dann der Fall, wenn sich aus Satzbau, Satzsinn und Zusammenhang des Textes erkennen lässt, dass der betreffende Satz direkt von einem Ergebnis, oder einer Schlussfolgerung zu einem Ergebnis, handelt.

Was gilt als Übereinstimmung:

Schlüsselpunkt Die Intervention:

- Entspricht die im Titel genannte Intervention der, in der Methodik als durchgeführt, beschriebenen so gilt dieses als Übereinstimmung.

- Eine Übereinstimmung wird auch dann angenommen, wenn die grundsätzliche Art und Weise der Intervention, wie im Titel beschrieben, mit derjenigen aus der Methodik übereinstimmt, obwohl in der Methodik weitergehende Definitionen der Intervention erfolgen, die im Titel nicht beschrieben sind. Dies bedeutete auch, dass in der Methodik ein zusätzlicher Interventionsarm definiert sein kann, solange dieser die aus dem Titel erwartete Intervention nicht nachhaltig verändert.

Schlüsselpunkt Das Ergebnis:

- Findet sich für das erwartete Ergebnis aus dem Titel ein beobachtetes Ergebnis in der Diskussion so gilt dieses als Übereinstimmung.
- Es gilt auch dann als Übereinstimmung, wenn mehr Ergebnisse in der Diskussion beobachtet wurden, als aus dem Titel zu erwarten gewesen wären, solange das erwartete Ergebnis aus dem Titel in der Diskussion beobachtet wurde.

Was gilt als mögliche Übereinstimmung:

Schlüsselpunkt Die Intervention:

- Wird die geplante Intervention im Titel nicht direkt genannt, kann aber aus dem Zusammenhang geschlossen werden, so gilt dies als mögliche Übereinstimmung, wenn diese so geschlossene Intervention der tatsächlich durchgeführten Intervention aus der Methodik im Grundsatz entspricht. Dies bedeutete, dass in der Methodik ein zusätzlicher Interventionsarm definiert sein kann, solange dieser die, aus dem Titel geschlossene, Intervention nicht nachhaltig verändert.
- Es gilt ebenfalls als mögliche Übereinstimmung, wenn nicht sicher festgestellt werden kann ob die geplante Intervention im Titel mit der durchgeführten Intervention in der Methodik übereinstimmt und gleichzeitig eine Nichtübereinstimmung nicht sicher ausgeschlossen werden kann.

Schlüsselpunkt Ergebnis:

- Kann nicht sicher entschieden werden ob für ein, aus dem Titel erwartetes, Ergebnis ein entsprechendes Ergebnis in der Diskussion beobachtet wurde und kann dieses ebenso nicht sicher ausgeschlossen werden, so gilt dieser Fall als mögliche Übereinstimmung.

Was gilt als Nichtübereinstimmung:

Schlüsselpunkt Die Intervention:

- Entspricht die als geplante Intervention im Titel genannte Intervention nicht im Grundsatz derjenigen in der Methodik beschriebenen, so gilt dies als Nichtübereinstimmung. Die geplante Intervention im Titel unterscheidet sich im Grundsatz von der tatsächlich durchgeführten Intervention in der Methodik dann, wenn durch einen zusätzlichen Interventionsarm in der Methodik die Intervention so nachhaltig verändert wird, dass sie sich deutlich von jener im Titel genannten Intervention unterscheidet.
- Es gilt ebenfalls als Nichtübereinstimmung, wenn sich für die im Titel definierte Intervention keine entsprechende Intervention in der Methodik wieder findet.

Schlüsselpunkt Ergebnis:

- Es gilt als Nichtübereinstimmung wenn sich für das aus dem Titel erwartete oder aus dem Titel sicher geschlossene Ergebnis kein beobachtetes Ergebnis in der Diskussion wieder findet.

<u>Für die vierte Ebene –Der primäre Endpunkt im Abstract- wird definiert:</u>

Identifikation des primären Endpunktes

Der primäre Endpunkt einer Studie ist derjenige Endpunkt, welcher explizit durch den Methodikteil als primärer Endpunkt benannt wird. Es gilt folglich nur derjenige Endpunkt als der primäre Endpunkt, der als solcher exakt definiert wurde, und nicht derjenige Endpunkt, welcher aus dem Zusammenhang der Studie heraus als primärer Endpunkt zu sehen ist.

.

Was gilt als direkte Bezugnahme:

- Beantwortet die Schlussfolgerung im Abstract direkt die, sich aus dem primären Endpunkt ergebende, Fragestellung, so gilt dieses als direkte Bezugnahme

Was gilt als mögliche Bezugnahme:

- Beantwortet die Schlussfolgerung im Abstract die Fragestellung des primären Endpunktes nicht direkt, so gilt es als mögliche Übereinstimmung, wenn die Schlussfolgerung im Abstract die Fragestellung für den primären Endpunkt indirekt beantwortet. Das bedeutet, dass die Schlussfolgerung das beobachtete Ergebnis für den primären Endpunkt nicht nennt, die in der Schlussfolgerung des Abstract getroffene Schlussfolgerung jedoch aufgrund des beobachteten Ergebnisses für den primären Endpunkt zulässig ist.

Was gilt als fehlende Bezugnahme:

- Bezieht sich die Schlussfolgerung in der Schlussfolgerung im Abstract auf das Ergebnis für einen Endpunkt, der nicht dem primären Endpunkt entspricht, so gilt dies als fehlende Bezugnahme auf den primären Endpunkt.
- Es gilt ebenso als fehlende Bezugnahme, wenn in der Methodik ein primärer Endpunkt nicht genannt oder nicht explizit als primärer Endpunkt definiert wurde.

Der Auswertungsbogen für die ICCT:

Studiencode	**Journal**	**Studientitel**	**Bearbeitungsdatum**
Jeder Studie wurde ein Code zugeordnet *Dieser beinhaltet ein Kürzel für das Journal und eine fortlaufende Nummer*	*Hier wurde der volle Name des Journals eingetragen*	*Dieses Feld enthielt den Originaltitel der bearbeiteten Studie*	*Das Datum, an dem diese Studie untersucht wurde*

Die erste Ebene:

E1	Textstelle Einleitung	Textstelle Methodik	Übereinstimmung (ja/nein/möglich)
Schlüsselpunkt Studiengruppe	*Studiengruppe aus der Einleitung*	*Studiengruppe aus der Methodik*	
Schlüsselpunkt Intervention	*Die Intervention aus der Einleitung*	*Die Intervention aus der Methodik*	
Schlüsselpunkt Das Ziel der Intervention	*Das Ziel der Intervention aus der Einleitung*	*Die definierten Endpunkte aus der Methodik*	

	Nichtübereinstimmung
Studiengruppe	
Die Intervention	
Das Ziel der Intervention	

Die zweite Ebene

E2A	Textstelle Methodik		Textstelle Diskussion	Übereinstimmung (ja/nein/möglich)
Studiengruppe	*Studiengruppe aus der Methodik*	**Studiengruppe**	*Studiengruppe aus der Diskussion*	
Das Ergebnis	*Endpunkte aus der Methodik*	**Das Ergebnis**	*beobachtete Ergebnisse in der Diskussion*	

E2B	Textstelle Diskussion	Textstelle Abstract	Übereinstimmung (ja/nein/möglich)
Studiengruppe	*Studiengruppe in der Diskussion*	*Studiengruppe im Abstract*	
Das Ergebnis	*beobachtetes Ergebnis in der Diskussion*	*Ergebnis aus dem Abstract*	

	Nichtübereinstimmung
E2A Studiengruppe	
E2A ERGEBNIS	
E2B Studiengruppe	
E2B ERGEBNIS	

Die dritte Ebene

E3A	Textstelle Titel	Textstelle Methodik	Übereinstimmung (ja/nein/möglich)
Intervention	*erwartete Intervention aus dem Titel*	*Durchgeführte Intervention in der Methodik*	

E3B	Textstellte Titel	Textstelle Diskussion	Übereinstimmung (ja/nein/möglich)
Ergebnis	*erwartetes Ergebnis aus dem Titel*	*beobachtetes Ergebnis in der Diskussion*	

	Nichtübereinstimmung
INTERVENTION E3A	
ERGEBNIS E3B	

Ebene	Übereinstimmung	Mögliche Übereinstimmung	Keine Übereinstimmung
E1			
E2A			
E2B			
E3A			
E3B			

Ebene	Mindestens eine Nichtübereinstimmung
E1	
E2	
E3	

8. Danksagung

Ich möchte mich ganz herzlich bei Herrn Prof. Dr. med. Franz Porzsolt für seine Geduld und Hilfsbereitschaft bedanken. Wann immer ein Problem zur Lösung anstand, konnte man stets mit seiner schnellen und kompetenten Hilfe rechnen. Ohne ihn wäre diese Arbeit nicht möglich gewesen.

Auch gilt mein besonderer Dank meinem alten Studienfreund Martin Gebhard, mittlerweile wie ich Assistenzarzt, ohne dessen Hilfe, Unterstützung und Motivation ich nicht zu dem Punkt gekommen wäre, an dem ich heute stehe.

Ich danke allen, die mich während der Arbeit an dieser Dissertation immer wieder unterstützt und motiviert haben. Besonders danke ich meinen Eltern ohne die mein Studium und somit auch diese Dissertation nicht möglich gewesen wären.

Zuletzt danke ich noch meiner Frau Karin für die Geduld die sie mir entgegen brachte und die Zeit die sie mir gab, diese Arbeit zu einem Ende zu bringen.

9. Lebenslauf

Persönliche Daten

Name:	Marc
Vorname:	Nagel
Geb. Datum:	20.07.1980
Geburtsort:	Überlingen
Anschrift:	Bergäcker 18 78355 Hohenfels
Familienstand:	verheiratet

Schulbildung:

1987-1991	Besuch der Grundschule in Hohenfels
1991-2000	Schüler am Nellenburg Gymnasium in Stockach
2000	Abschluss mit der Allgemeinen Hochschulreife

Zivildienst:

2000-2001	Zivildienstleistender im Rettungsdienst

Hochschulbildung:

2001-2006	Studium der Humanmedizin an der Universität Ulm
20.09.2003	Ärztliche Vorprüfung (Note „gut“/2,0)
August 2006 bis Juli 2007	Praktisches Jahr am Klinikum Kempten Oberallgäu
09.05.2008	Abschluss der Ärztlichen Prüfung (Note „gut“/2,0)
19.05.2008	Approbation als Arzt

Beruflicher Werdegang:

Seit Juli 2008	Assistenzarzt für Chirurgie am Städtischen Krankenhaus in Stockach