

Universitätsklinikum Ulm
Zentrum für Chirurgie
Klinik für Allgemein- und Viszeralchirurgie
Ärztlicher Direktor: Prof. Dr. med. Christoph Michalski

**Validität präoperativer Chemotherapie für das resektable
thorakale Ösophaguskarzinom**
Eine kritische Bewertung randomisierter Studien

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

Ursula Marianne Alexandra Klotz

Feldkirch

2021

Amtierender Dekan: Prof. Dr. Thomas Wirth

1. Berichterstatter: PD Dr. med. Michael Kremer

2. Berichterstatter: Prof. Dr. med. Johannes Weiss

Tag der Promotion: 28.10.2022

Meinen Eltern in Liebe und Dankbarkeit

Teile dieser Dissertation wurden bereits in folgendem Fachartikel veröffentlicht:

Manzini G, Klotz U, Henne-Bruns D, Kremer M:

Validity of studies suggesting preoperative chemotherapy for resectable thoracic esophageal cancer: A critical appraisal of randomized trials.

World Journal of Gastrointestinal Oncology 12 (2020)

DOI: 10.4251/wjgo.v12.i1.113

Inhaltsverzeichnis

INHALTSVERZEICHNIS.....	I
ABKÜRZUNGSVERZEICHNIS	III
1 EINLEITUNG	1
2 MATERIAL UND METHODEN	6
2.1 Cochrane Review <i>Kidane et al.</i>	6
2.1.1 Metaanalyse: Gesamtüberleben	6
2.2 CONSORT Statement	8
2.3 Auswahl der Studien	14
2.4 Neuberechnete Metaanalysen	15
3 ERGEBNISSE	21
3.1 Validitätsbewertung der signifikanten Studien	21
3.1.1 Ychou et al. 2011	24
3.1.2 MRC Allum et al. 2009	25
3.1.3 Boonstra et al. 2011	27
3.1.4 Law et al. 1997	27
3.2 Validitätsbewertung der nicht-signifikanten Studien	28
3.3 Ergebnisse der neuen Metaanalysen.....	33
3.3.1 Metaanalyse I	33
3.3.2 Metaanalyse II	34
4 DISKUSSION.....	36
4.1 Randomisation, Stratifikation und Allocation Concealment	37
4.2 Verblindung und Placebo-Kontrolle.....	38
4.3 Fallzahlberechnung	39
4.4 Intention-to-treat-Analyse	40
4.5 Ort der Intervention.....	40

4.6	Interessenskonflikte.....	41
4.7	Limitationen	42
4.8	Schlussfolgerungen	43
5	ZUSAMMENFASSUNG	46
6	LITERATURVERZEICHNIS.....	48
	DANKSAGUNGEN	58
	LEBENS LAUF	59

Abkürzungsverzeichnis

5-FU	5-Fluoruracil
CI	Konfidenzintervall
CTx	Chemotherapie
Gy	Gray
HR	Hazard Ratio (Überlebenszeitanalyse)
I^2	I-Quadrat
i.v.	intravenös
n	Patienten-/Probandenanzahl
N(C)	Patientenanzahl in Kontrollgruppe
N(T)	Patientenanzahl in Interventionsgruppe
No.	Nummer
OP	Operation
p.o.	per os
präop.	präoperativ
R0	Kein Residualtumor
R1	Mikroskopischer Residualtumor
R2	Makroskopischer Residualtumor
RCTx	Radiochemotherapie
RTx	Radiotherapie
S.	Seite
SE	Standardfehler
T N M	Tumor Lymphknoten Metastasen (gemäß TNM-Klassifikation)
W(fixed)	durch Festeffektmodell zugewiesenes Gewicht
W(random)	durch Zufallseffektmodell zugewiesenes Gewicht
WHO	Weltgesundheitsorganisation
τ^2	Tau-Quadrat
χ^2	Chi-Quadrat

1 Einleitung

Das Ösophaguskarzinom ist die achthäufigste Tumorerkrankung des Menschen und die sechshäufigste tumorbedingte Todesursache [23]. Mit im Jahr 2020 604.000 Neuerkrankungen weltweit wurden 3,1 % aller Krebserkrankungen durch diese Tumorart bedingt [75]. Es zählt mit einer relativen 10-Jahres Überlebensrate von 21 % bei Frauen beziehungsweise 17 % bei Männern zu den Tumorarten mit einer sehr schlechten Prognose [59]. Bei einem Großteil der Patienten liegt, verbunden mit einer späten Diagnosestellung, bereits ein lokal fortgeschrittenes Stadium vor.

Das lokal fortgeschrittene Stadium ist definiert als eine auf den Ösophagus oder das resezierbare periösophageale Gewebe beschränkte Tumorausdehnung (entspricht T2-T4 gemäß TNM-Klassifikation) und/oder einem lokalen Lymphknotenbefall (N1-N3) ohne Fernmetastasierung (M0) [32].

Die einzige potenziell kurative Behandlungsform in diesem Stadium ist seit ihrer Entwicklung zu Beginn des 20. Jahrhunderts die chirurgische Resektion des Ösophagus und stellt somit den Goldstandard der Therapie dar [39]. Eine Ösophagusresektion in kurativer Absicht ist jedoch nur in 15 % bis 39 % der Fälle möglich [38; 43; 40; 54].

Zudem ist diese Therapiemodalität – alleinig angewandt – aufgrund von frühzeitig auftretenden okkulten Mikrometastasen häufig nicht in der Lage, den natürlichen Krankheitsverlauf aufzuhalten. Betroffene versterben im Erkrankungsverlauf sowohl an Lokalrezidiven als auch an Fernmetastasen [32]. Die mediane Überlebenszeit nach Resektion in kurativer Absicht beträgt somit lediglich 15 bis 18 Monate mit einer 5-Jahres Überlebensrate von 20 % bis 25 % [58].

Hohen Rezidiv- und geringen Überlebensraten geschuldet galt es in den letzten drei Jahrzehnten, unter der Zielsetzung eines verbesserten Gesamtüberlebens, die Rolle zusätzlicher Therapiemodalitäten in einem multidisziplinären Therapieregime zu definieren [32]. Die optimale Kombination aus Operation, Chemotherapie, Radiotherapie oder kombinierte Radiochemotherapie und der günstigste Zeitpunkt ihrer Implementierung, prä-, peri- oder postoperativ, wird nach wie vor kontrovers diskutiert. Obwohl zahlreiche randomisiert-kontrollierte Studien und Metaanalysen versucht haben, sich dieser Problematik anzunähern, unterscheiden sich internationale

Leitlinien erheblich oder geben keine klare Empfehlung für ein spezifisches Behandlungsregime ab [24]. Eine dieser zusätzlichen Therapiemodalitäten ist eine präoperative (Radio-) Chemotherapie, die darauf abzielt, die Mikrometastasenlast zu reduzieren, die Resektabilität des Primärtumors zu verbessern und Symptome wie Dysphagie zu lindern [24; 42].

Klinische Studien lieferten bislang keine einheitlichen Resultate zur Verbesserung des Gesamtüberlebens durch eine präoperative Chemotherapie. So zeigte ein erster Cochrane Review von *Malthaner et al.* aus dem Jahr 2001 keinen Überlebensvorteil für Patienten nach präoperativer Chemotherapie bei resektablem Ösophaguskarzinom [47]. Die Metaanalyse von *Urschel et al.* aus dem Jahr 2002 konnte nach Einschluss von elf randomisierten Studien ebenfalls keinen Überlebensvorteil feststellen [80].

Erst in den Studien von *Boonstra et al.* (2011) [10], *MRC Allum et al.* (2009) [1] und *Ychou et al.* (2011) [84] konnte ein signifikanter Überlebensvorteil für Patienten mit präoperativer Chemotherapie festgestellt werden. Nach Einschluss dieser Studien in die Metaanalyse eines im Jahr 2015 veröffentlichten Cochrane Reviews von *Kidane et al.* zeigte sich eine signifikante Verbesserung des Gesamtüberlebens [Hazard Ratio (HR): 0,88; 95 %- Konfidenzintervall (CI): 0,80 bis 0,96, Abbildung 1] [34]. Insgesamt wurden zehn randomisiert-kontrollierte Studien mit dem Gesamtüberleben als primären Endpunkt in diese Metaanalyse eingeschlossen.

Randomisiert-kontrollierte Studien und Systematische Reviews:

Ein hochwertiges Studiendesign und der Goldstandard für die Erforschung von Wirksamkeit und Sicherheit neuer und bestehender Therapieansätze sind randomisiert-kontrollierte Studien [31]. Sie weisen bei korrekter Durchführung die höchste interne Validität auf.

Hierbei ist die interne Validität ein Gütekriterium, das Aufschluss darüber gibt, in welchem Maße erhobene Messwerte mit den im Vorfeld festgelegten Messwerten übereinstimmen und damit dem gemessenen Effekt vertraut werden kann. Sie wird von der Durchführung, Auswertung und Berichterstattung einer Studie beeinflusst [30]. Messungen sind demnach dann valide, wenn die Forschungsfrage einer Studie mit den angewandten Messmethoden und den gemessenen Parametern tatsächlich beantwortet werden kann.

Um dies zu erreichen, müssen Studien einem Aufbau folgen, der Störvariablen (sogenannte *confounder*) und systematische Fehler (sogenannter *bias*) auf ein Minimum reduziert. Die Hierarchie eines Studiendesigns erstreckt sich mit abnehmender Fehleranfälligkeit somit von Expertenmeinungen, Fallberichten, Fall-Kontroll-Studien, Kohorten-Studien hin zu randomisiert-kontrollierten Studien, wobei letztere aufgrund ihres Designs das geringste Verzerrungsrisiko aufweisen.

Demgegenüber steht die externe Validität, die eine Aussage darüber erlaubt, ob sich die Studienergebnisse auch auf andere Studienkohorten übertragen lassen. Sie spiegelt die Generalisierbarkeit der Resultate wider und wird von der Thematik, den Ein- und Ausschlusskriterien und dem Setting einer Studie bestimmt [30].

Die fortlaufende Generierung neuer Studien führt einerseits zu einem bedeutsamen Wissenszuwachs in der Humanmedizin, andererseits jedoch zu einer großen Unübersichtlichkeit und teils widersprüchlichen Forschungsergebnissen, die es den Behandlern erschweren, die für eine klinische Fragestellung relevanten Studien zu identifizieren und zu bewerten. Aus diesem Grund werden Übersichtsarbeiten, sogenannte *systematische Reviews*, angefertigt, die Ergebnisse relevanter Primärstudien mit der gleichen Fragestellung deskriptiv zusammenfassen und beurteilen [15]. *Systematisch* bedeutet in diesem Zusammenhang eine nach festgelegten Kriterien reproduzierbare und methodische Vorgehensweise und unter der Kontrolle von Bias. Ziel ist also eine Verdichtung und kritische Bewertung von aus Einzelstudien gewonnenen Informationen.

Liegen die Ergebnisse in einer systematisierten Form vor, wird eine Übersichtsarbeit häufig um eine Metaanalyse ergänzt. Diese arbeitet die Ergebnisse empirischer Primärstudien zur Effektivität einer Intervention zusätzlich mittels quantitativ statistischer Methoden auf und fasst diese zu einem globalen Ergebnis zusammen [12]. Der wesentliche Ansatz einer Metaanalyse besteht in der Erklärung der Variabilität zwischen den Ergebnissen einzelner Primärstudien, die häufiger keine eindeutigen oder sogar gegensätzliche Schlussfolgerungen hinsichtlich einer Fragestellung ziehen.

Durch die Kombination der in den einzelnen Primärstudien gemessenen Effektgrößen, und unter Berücksichtigung der Studienfallzahlen, wird ein Gesamteffekt einer bestimmten Intervention geschätzt. Diese Effektgröße (auch *Effektschätzer* oder

Therapieeffekt genannt) ist ein Maß, das angibt, wie stark sich zwei Gruppen, beispielsweise eine Interventions- und Kontrollgruppe, voneinander unterscheiden. Dadurch lassen sich die Ergebnisse verschiedener Primärstudien anhand eines einheitlichen Maßes miteinander vergleichen.

Als Sekundäranalyse bezieht eine Metaanalyse somit ihre Daten aus den Ergebnissen bereits vorliegender Untersuchungen, um einen Überblick zu geben, welche Forschung zu einer Fragestellung bereits existiert, neue Fragen zu beantworten und um allgemeingültigere Aussagen mit dem Ziel einer evidenzbasierten Konsensbildung zu schaffen [15; 25].

Cochrane Collaboration

Die seit 1993 bestehende *Cochrane Collaboration* ist ein unabhängiges Netzwerk aus Wissenschaftlern, Ärzten, Gesundheitsfachpersonal und Patienten aus über 130 Staaten. Ihre Zielsetzung ist die Erstellung, Aktualisierung und Verbreitung hochwertiger systematischer Reviews basierend auf systematischer Literatursuche, Qualitätsbewertung der eingeschlossenen Studien, quantitativer Ergebnisanalyse und objektiver Interpretation der Daten aus systematischen Übersichtsarbeiten, um medizinische Information auf Basis bestmöglicher empirischer Evidenz zu Verfügung stellen [16]. Evidenzbasierte Medizin vereint bei medizinischen Fragestellungen und Therapieentscheidungen Expertenmeinung und Patientenwunsch mit wissenschaftlich nachgewiesener Wirksamkeit [62]. Randomisiert-kontrollierte Studien und deren Aufarbeitung in systematischen Reviews mit Metaanalysen sind somit eine wesentliche Komponente für den Transfer wissenschaftlicher Forschungsergebnisse in die klinische Praxis.

CONSORT Statement 2010

Das *CONSORT Statement 2010* ist eine Leitlinie für die Verfassung von Forschungsberichten (sogenannte *Reporting Guideline*). Unvollständige, unscharfe und unverständliche Studienbeschreibungen führten im Jahr 1996 mit dem CONSORT Statement erstmals zur Entwicklung einer solchen Reporting Guideline für randomisiert-kontrollierte Studien [7]. Im Jahr 2001 wurde das CONSORT Statement ins Deutsche übersetzt [52] und letztmalig im Jahr 2010 aktualisiert [66]. In Form einer Checkliste mit 37 Kriterien und eines Flussdiagrammes werden Empfehlungen

für die wissenschaftlich adäquate Publikation von randomisiert-kontrollierten Studien abgegeben [55]. Sie dient Studienautoren, Studienberichte transparent und vollständig zu verfassen.

Darüber hinaus benötigt der Leser für eine valide und genaue Beurteilung einer Studie eine qualitativ hochwertige Berichterstattung mit genügender Transparenz und genauen Informationen zu der angewandten Methodik und den erhobenen Ergebnissen. Reporting Guidelines stellen somit eine Möglichkeit der Orientierung dar, die einen Überblick über die methodische Qualität einer Studie bietet [50].

Hohe methodische Anforderungen an klinische Studien und eine genügende Publikationsqualität bieten Schutz vor systematischen Verzerrungen und können so zur Vermehrung von validem, evidenzbasierten und medizinischen Wissen beitragen. Da Cochrane Reviews und deren Metaanalysen zu den qualitativ hochwertigsten Publikationen gehören, werden deren Empfehlungen häufig in Leitlinien zu Diagnostik und Therapie unterschiedlichster Erkrankungen aufgenommen.

Fragestellung

Aufgrund des hohen Stellenwertes die Cochrane Reviews in der Gewinnung evidenzbasierter Therapieempfehlungen besitzen und ihrer häufigen Einbindung in Leitlinien, sind deren Erkenntnisse insbesondere auf hinreichende Validitätskriterien zu prüfen.

Ziel dieser Arbeit ist die Untersuchung der Validität der Metaanalyse des Cochrane Reviews von *Kidane et al.* von 2015 anhand einer Validitätsbewertung der zugrundeliegenden Einzelstudien. Insbesondere soll hierbei die Validität der signifikant für eine präoperative Chemotherapie sprechenden Studien geprüft und anschließend die Aussagekraft der Metaanalyse erneut bewertet werden.

2 Material und Methoden

2.1 Cochrane Review *Kidane et al.*

Der Cochrane Review von *Kidane et al.* aus dem Jahr 2015 basiert auf 13 randomisiert-kontrollierten Studien, die insgesamt 2362 Patienten einschlossen. Die Rolle einer präoperativer Chemotherapie bei resektablem thorakalen Ösophaguskarzinom sollte untersucht werden [34].

Als primären Studienendpunkt wurde seitens der Autoren das Gesamtüberleben festgelegt. Weitere definierte Endpunkte waren die Resektionsrate, das Auftreten von Tumorrezidiven, das Ansprechen auf die Chemotherapie, die Behandlungsmorbidität und -mortalität sowie der Zugewinn an Lebensqualität [34].

Eingeschlossen wurden Studien, die Patienten mit potenziell resezierbarem, thorakalem Ösophaguskarzinom (unabhängig des histopathologischen Subtyps) einschlossen. Nach dem Zufallsprinzip wurden die zwei Gruppen, *präoperative Chemotherapie mit anschließender Operation* oder *alleinige Operation*, gebildet. Ausgeschlossen wurden nicht-randomisierte Studien, Studien, die mit anderen Therapiemodalitäten (beispielsweise: Radiotherapie) behandelte Patienten einschlossen, oder Studien mit Patienten mit zervikalem Ösophaguskarzinom. Anhand des Datensatzes dieser 13 Studien führten die Autoren des Reviews insgesamt 18 Metaanalysen mit den oben genannten Endpunkten durch [34].

Für die in dieser Arbeit durchgeführte Validitätsbewertung der Fragestellung eines verbesserten Gesamtüberlebens mit präoperativer Chemotherapie wurde die Metaanalyse mit dem Endpunkt *Gesamtüberleben* des Cochrane Reviews betrachtet.

2.1.1 Metaanalyse: Gesamtüberleben

In die Metaanalyse des Cochrane Reviews von *Kidane et al.* mit dem primären Endpunkt *Gesamtüberleben* wurden zehn Studien mit insgesamt 2122 Patienten eingeschlossen. Drei der insgesamt 13 Studien des Cochrane Reviews, *Baba et al.* (2000) [4], *Stilidi et al.* (2006) [74] und *Wang et al.* (2001) [82], enthielten ungenügende Informationen zur Berechnung einer Hazard Ratio und wurden seitens der Studienautoren aus der Metaanalyse ausgeschlossen.

Vier der zehn Studien (40 %; *Boonstra et al.* (2011) [10], *MRC Allum et al.* (2009) [1], *Law et al.* (1997) [37] und *Ychou et al.* (2011) [84]) zeigten einen statistisch signifikanten Überlebensvorteil für Patienten mit präoperativer Chemotherapie.

Die anderen sechs Studien (60 %) zeigten kein statistisch signifikantes Ergebnis für eine präoperative Chemotherapie (*Ancona et al.* (2001) [2], *Kelsen et al.* (1998) [33], *Maipang et al.* (1994) [46], *Nygaard et al.* (1992) [53], *Roth et al.* (1988) [61] und *Schlag et al.* (1992) [63]).

Kidane et al. kam mit dem Ergebnis ihrer Metaanalyse zu der Schlussfolgerung, dass eine präoperative Chemotherapie mit nachfolgender Operation bei resektablem thorakalen Ösophaguskarzinom das Gesamtüberleben der Patienten statistisch signifikant verbessert (HR: 0,88; 95 %-CI: 0,80 bis 0,96, Abbildung 1) [34].

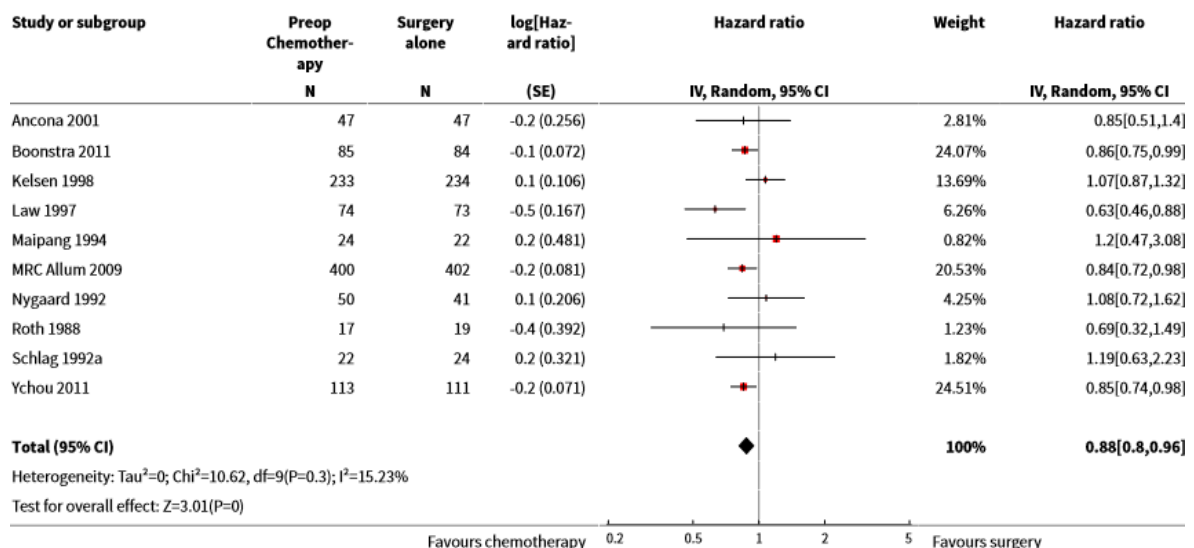


Abbildung 1: Metaanalyse: Gesamtüberleben des Cochrane Reviews: Links stehen die zehn Studien mit der jeweiligen Patientenanzahl in der Interventions- und Kontrollgruppe mit Angabe des Standardfehlers und der logarithmischen Hazard Ratio. Mittig wird die Hazard Ratio als rotes Quadrat und das 95 %-Konfidenzintervall als waagerechte Verlängerung graphisch in einem Forest-Plot dargestellt. Die schwarze Raute symbolisiert die Gesamteffektgröße. Rechts werden die Gewichtungen der Einzelstudien prozentual und die Hazard Ratio als 95 %-Konfidenzintervall numerisch angegeben. Die Heterogenität wird entsprechend ihrer Kennwerte Tau², Chi², I² und p angegeben.
Abkürzungen: CI: Konfidenzintervall, df: degrees of freedom, N: Patientenanzahl, SE: Standardfehler, P: P-Wert, IV: Intervall, Z: Z-Wert

Quelle: Abbildung 1 stammt aus Kidane B, Coughlin S, Vogt K, Malthaner R: Preoperative chemotherapy for resectable thoracic esophageal cancer. *Cochrane Database of Systematic Reviews* 5: S. 36 (2015) [34], DOI: 10.1002/14651858.CD001556.pub3.

Copyright © 2015 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

2.2 CONSORT Statement

Das *CONSORT* Statement 2010 (Consolidated Standards Of Reporting Trials) bietet Studienautoren anhand einer Checkliste aus 37 Kriterien (Tabelle 2) und einem Flussdiagramm (Abbildung 2) eine standardisierte Vorlage zur Berichterstattung über randomisiert-kontrollierten Studien. Die 37 Kriterien sind sieben Themenbereichen zugeordnet [66]:

1. Titel und Zusammenfassung
2. Einleitung
3. Methoden
4. Randomisierung
5. Ergebnisse
6. Diskussion
7. Andere Informationen

Diese Leitlinie beinhaltet evidenzbasierte Mindeststandards für die Durchführung und Dokumentation randomisiert-kontrollierter Studien im Parallelgruppen-Design, mit dem Ziel deren Qualität, Transparenz und Plausibilität der Berichterstattung zu verbessern und damit auch eine wahrheitsgemäße Interpretation der Ergebnisse zu gewährleisten.

Für die in dieser Arbeit durchgeführte Validitätsbewertung eines verbesserten Gesamtüberlebens mit präoperativer Chemotherapie wurden die einzelnen Kriterien der *CONSORT* Checkliste als *Validitätskriterien* bezeichnet. Die zehn Studien der Metaanalyse *Gesamtüberleben* von *Kidane et al.* [34] wurden auf alle 37 Validitätskriterien hin geprüft und einem der drei nachfolgenden Ergebnisse zugeordnet:

1. Trifft das Validitätskriterium auf eine Studie zu, so gilt es als erfüllt und ein *Ja* wurde vergeben
2. Trifft das Validitätskriterium auf eine Studie nicht zu, so gilt es als nicht erfüllt und ein *Nein* wurde vergeben
3. Kann das Validitätskriterium nicht angewandt werden oder ist in einer Studie nicht von Relevanz, wurde ein *Nicht anwendbar* vergeben

Die einzelnen Validitätskriterien werden im Folgenden und in Tabelle 2 näher erläutert [51]:

1a: Es soll aus dem Titel einer Studie eindeutig hervorgehen, dass es sich um eine *randomisierte* Studie handelt.

1b: Nachfolgend soll ein strukturiertes Abstrakt mit Zusammenfassung von Studiendesign, Methoden, Ergebnisse und Schlussfolgerungen folgen. Das CONSORT Statement 2010 stellt hierbei eine Liste essentieller Kriterien für das Abstract einer Studie bereit (Tabelle 1).

2a: Es soll der wissenschaftliche Hintergrund und die Gründe für die Durchführung der Studie erläutert werden.

2b: Es sollen genaue Fragestellungen oder Hypothesen aufgestellt werden.

3a: Es soll das genaue Studiendesign (beispielsweise: parallel, crossover, cluster oder faktoriell) und das Zuordnungsverhältnis explizit angegeben werden. Bei Arzneimittelstudien ist die Phase der Studie (I-IV) relevant.

3b: Alle Änderungen der Methoden nach Studienbeginn sollen mit Gründen angegeben werden. Unabhängig davon, ob die Änderungen ausdrücklich Teil des Studiendesigns sind oder eine Reaktion auf sich ändernde Umstände darstellen.

4a: Die genauen Eignungskriterien der Probanden sollen angegeben werden, um die Generalisierbarkeit und Relevanz der Studie einschätzen zu können.

4b: Umgebung und Ort der Studiendurchführung soll angegeben werden.

5: Die Autoren sollen jede Intervention einschließlich Kontrollinterventionen gründlich beschreiben. Es soll ermöglicht werden, die Intervention wiederholen zu können und genau zu wissen, wie die Intervention durchzuführen ist.

Bei Arzneimittelintervention sollen die Informationen dabei den Namen, Dosis, die Art der Verabreichung des Medikaments (beispielsweise: p.o., i.v.), Zeitpunkt und Dauer der Verabreichung beinhalten, Bedingungen, unter denen Interventionen zurückgehalten, und, falls zutreffend, Titrationsschema angegeben werden.

Tabelle 1: Liste essentieller Kriterien eines Abstracts einer randomisiert-kontrollierten Studie
Quelle: Tabelle 1 stammt aus Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG: CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *British Medical Journal (BMJ)* 340: S. 4 (2010) [51], DOI: <https://doi.org/10.1136/bmj.c869> und wurde unter CC BY-NC 4.0 (<http://creativecommons.org/licenses/by-nc/4.0/>) publiziert.

Item	Description
Authors	Contact details for the corresponding author
Trial design	Description of the trial design (such as parallel, cluster, non-inferiority)
Methods:	
Participants	Eligibility criteria for participants and the settings where the data were collected
Interventions	Interventions intended for each group
Objective	Specific objective or hypothesis
Outcome	Clearly defined primary outcome for this report
Randomisation	How participants were allocated to interventions
Blinding (masking)	Whether participants, care givers, and those assessing the outcomes were blinded to group assignment
Results:	
Numbers randomised	Number of participants randomised to each group
Recruitment	Trial status
Numbers analysed	Number of participants analysed in each group
Outcome	For the primary outcome, a result for each group and the estimated effect size and its precision
Harms	Important adverse events or side effects
Conclusions	General interpretation of the results
Trial registration	Registration number and name of trial register
Funding	Source of funding

6a: Randomisiert-kontrollierte Studien bewerten Ergebnisse beziehungsweise Endpunkte, für die die Gruppen verglichen werden. Meistens haben Studien mehrere Endpunkte. Es sollen somit alle primären und sekundären Endpunkte angegeben werden, einschließlich wie und wann sie erhoben wurden.

6b: Wenn vorhanden, sollen alle Änderungen der Endpunkte nach Studienbeginn unter Angabe von Gründen aufgeführt werden.

7a: Aus wissenschaftlichen und ethischen Gründen muss die Stichprobengröße für eine Studie sorgfältig geplant werden, wobei ein Gleichgewicht zwischen klinischen und statistischen Überlegungen zu berücksichtigen ist. Es soll folgedessen genau angegeben werden, wie die Fallzahl berechnet wurde.

7b: Falls zutreffend, sollen alle Zwischenanalysen und Abbruchkriterien erklärt werden.

8a: Die Probanden beziehungsweise Patienten sollten auf der Grundlage eines zufälligen Prozesses, der durch Unvorhersehbarkeit gekennzeichnet ist, in Vergleichsgruppen zugeordnet werden. Die Autoren sollen ausreichende Informationen bereitstellen, damit der Leser die Methoden zur Erzeugung der zufälligen Zuordnungssequenz und die Wahrscheinlichkeit einer Verzerrung bei der Gruppenzuweisung beurteilen kann. Es soll die Methode zur Generierung der Zufallszuteilung angegeben werden.

8b: Es soll die genaue Art der Randomisierung angegeben werden (beispielsweise: Blockrandomisierung, Stratifikation, Minimisation)

9: In Punkt *8a* wurde die Erzeugung einer unvorhersehbaren Folge von Zuweisungen verdeutlicht. Studienautoren sollen unter Punkt 9 nun beschreiben, wie sie eine angemessene Geheimhaltung dieser Randomisierungsliste implementiert haben.

10: Es soll beschrieben werden, wer die Zufallszuordnung durchführte, wer die Teilnehmer in die Studie aufnahm und wer die Teilnehmer den Interventionen zuteilte.

11a: Es soll beschrieben werden, wer von den Beteiligten verblindet war.

11b: Die Autoren sollen die Ähnlichkeit der Interventionen beschreiben (beispielsweise: Aussehen, Geschmack, Farbe).

12a: Es soll genau beschrieben werden, welche statistischen Methoden für den Vergleich der Gruppen hinsichtlich ihrer primären und sekundären Endpunkte verwendet wurden.

12b: Falls relevant, soll beschrieben werden, welche statistischen Methoden für weitere Analysen, wie Subgruppenanalysen, verwendet wurden.

13a: Es soll die genaue Anzahl der Probanden, die zugeteilt, die geplante Intervention erhielten und die hinsichtlich des primären Endpunktes analysiert wurden, angegeben werden.

13b: Falls relevant, soll die genaue Anzahl nach Randomisation ausgeschlossener Probanden mit Gründen angegeben werden.

14a: Wie lange wurde rekrutiert und nachbeobachtet?

14b: Warum endete die Studie?

15: Es soll eine Tabelle mit klinischen und demographischen Charakteristika für jede Gruppe (Vergleichs- und Kontrollgruppen) eingefügt werden.

16: Für jede Gruppe soll die Anzahl der Teilnehmer, die in die Gruppe eingeschlossen wurden, angegeben werden und, ob diese der Gruppengröße der ursprünglich geplanten Gruppengröße entsprach.

17a: Es sollen für jeden primären und sekundären Endpunkt die Ergebnisse jeder Gruppe und deren geschätzte Effektgröße sowie ihre Präzision (beispielsweise das 95 %-Konfidenzintervall) angegeben werden.

17b: Es sollen sowohl absolute als auch relative Effektgrößen für binäre Endpunkte angegeben werden.

18: Die Ergebnisse aller weiteren Analysen sollen angegeben werden.

19: Die Autoren sollen alle möglichen unerwünschten Arzneimittelwirkungen und Komplikationen innerhalb der Gruppen beschreiben.

20: Es sollen mögliche Studienlimitierungen und fehlende Präzision angegeben werden.

21: Sind die Studienergebnisse generalisierbar? Ist die externe Validität gewährleistet?

22: Ist die Interpretation der Studienautoren vereinbar mit den Studienergebnissen?

23: Es sollen Registriernummer und Name des Registers angegeben werden.

24: Falls verfügbar, soll angegeben werden, wo das vollständige Studienprotokoll einsehbar ist.

25: Jegliche finanzielle Unterstützung und Rolle von Geldgebern und Sponsoren sollen angegeben werden.

Tabelle 2: CONSORT Checkliste 2010**Abkürzungen:** No: Number

Quelle: *Tabelle 2 stammt aus Schulz K F, Altman D G, Moher D, CONSORT Group: CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. PLoS Medicine 7: S. 2 (2010) [65], DOI: <https://doi.org/10.1371/journal.pmed.1000251> und wurde unter CC BY-NC 4.0 (<http://creativecommons.org/licenses/by-nc/4.0/>) publiziert.*

Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract			
	1a	Identification as a randomised trial in the title	
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts [21,31])	
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	
	2b	Specific objectives or hypotheses	
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	
Participants	4a	Eligibility criteria for participants	
	4b	Settings and locations where the data were collected	
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	
	6b	Any changes to trial outcomes after the trial commenced, with reasons	
Sample size	7a	How sample size was determined	
	7b	When applicable, explanation of any interim analyses and stopping guidelines	
Randomisation:			
Sequence generation	8a	Method used to generate the random allocation sequence	
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	
	11b	If relevant, description of the similarity of interventions	
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	
	13b	For each group, losses and exclusions after randomisation, together with reasons	
Recruitment	14a	Dates defining the periods of recruitment and follow-up	
	14b	Why the trial ended or was stopped	
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms [28])	
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	
Other information			
Registration	23	Registration number and name of trial registry	
Protocol	24	Where the full trial protocol can be accessed, if available	
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	

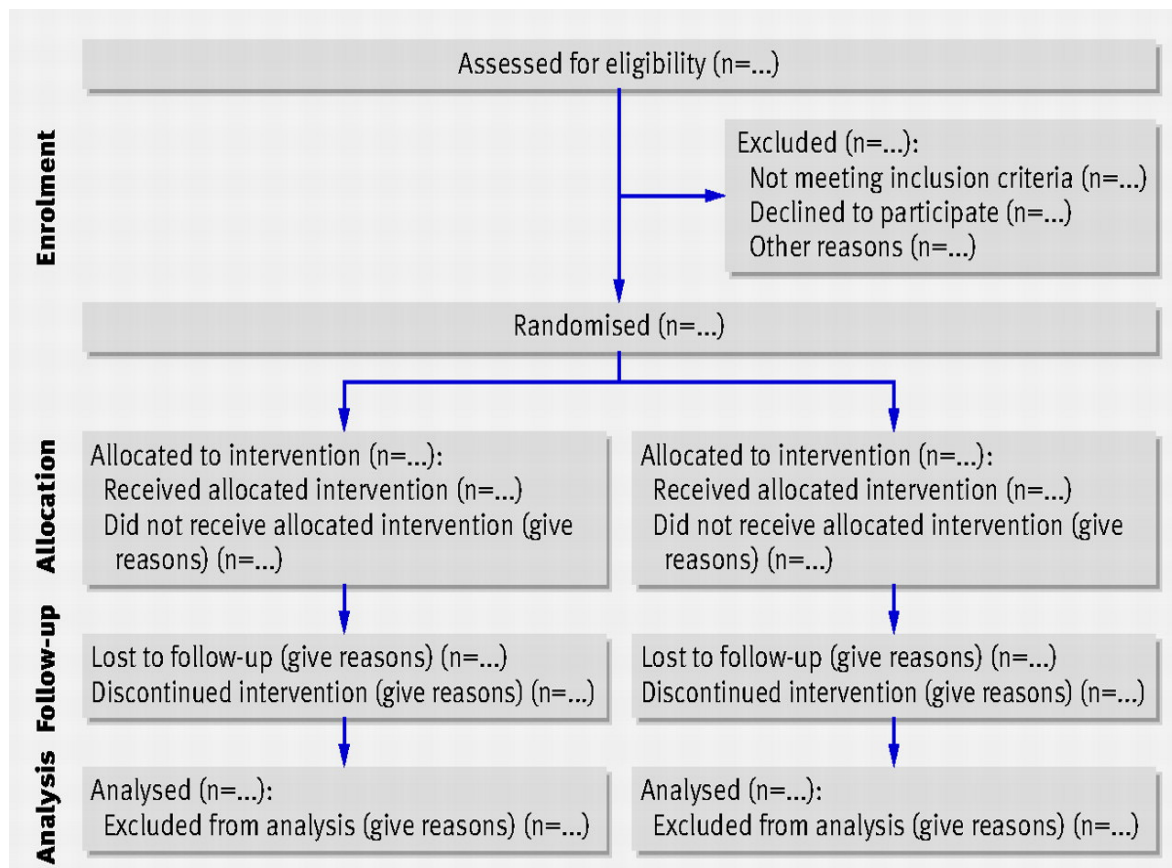


Abbildung 2: Flussdiagramm durch die Phasen einer parallel randomisierten Studie mit zwei Gruppen: Registrierung und Beurteilung der Eignung eines Probanden (*Enrolment*), Randomisierte Interventionszuweisung (*Allocation*), Nachbeobachtung (*Follow-up*) und Datenanalyse (*Analysis*).
Abkürzungen: *n*: Anzahl der jeweiligen Patienten
Quelle: Abbildung 4 stammt aus Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG: CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *British Medical Journal (BMJ)* 340: S. 2 (2010) [51], DOI: <https://doi.org/10.1136/bmj.c332> und wurde unter CC BY-NC 4.0 (<http://creativecommons.org/licenses/by-nc/4.0/>) publiziert.

2.3 Auswahl der Studien

Vier der in die Metaanalyse von *Kidane et al.* [34] eingeschlossenen Studien zeigten einen statistisch signifikanten Überlebensvorteil für Patienten mit präoperativer Chemotherapie (*Boonstra et al.* [10], *MRC Allum et al.* [1], *Law et al.* [37] und *Ychou et al.* [84]). Davon wurden die drei Studien mit der höchsten Gewichtung (*Boonstra et al.* [10], *MRC Allum et al.* [1] und *Ychou et al.* [84]) zur Neuberechnung der Metaanalysen ausgewählt.

Für diese Gewichtung ausschlaggebenden Parameter waren: *Stichprobengröße*, *Schätzgenauigkeit* und *Breite des Konfidenzintervalls*. Dabei wurde *Boonstra et al.* [10] mit 24,1 %, *MRC Allum et al.* [1] mit 20,5 % und *Ychou et al.* [84] mit 24,5 % gewichtet (Abbildung 1) und sind somit ausschlaggebend für die Empfehlung einer

präoperativen Chemotherapie [34]. Die Studie *Law et al.* [37] wurde im Vergleich zu den zuvor genannten Studien nur mit 6,3 % gewichtet [34].

2.4 Neuberechnete Metaanalysen

Zur Beantwortung der Fragestellung dieser Arbeit wurden zwei neue Metaanalysen durchgeführt:

1. Metaanalyse I erfolgte ohne die drei Studien mit höchster Gewichtung (*Boonstra et al.* [10], *MRC Allum et al.* [1] und *Ychou et al.* [84]) (n=7).
2. Metaanalyse II schloss die Studie mit der höchsten Validität gemäß CONSORT Statement 2010 (*Boonstra et al.* [10]) mit ein (n=8).

In beide neu berechneten Metaanalysen wurden die übrigen sieben Studien miteingeschlossen (*Ancona et al.* [2], *Kelsen et al.* [33], *Law et al.* [37], *Maipang et al.* [46], *Nygaard et al.* [53], *Roth et al.* [61] und *Schlag et al.* [63]).

Grundsätzlich sind für die Erstellung einer Metaanalyse zwei Schritte notwendig: Als erstes wird die Effektgröße definiert und für jede eingeschlossene Studie berechnet. Da in der Untersuchung die Gesamtüberlebensrate als primären Endpunkt ausgewählt wurde, wird die Effektgröße durch die Hazard Ratio als ideales numerisches Maß für den Vergleich von Überlebenszeiten beschrieben.

Hazard Ratio

Hazard (zu Deutsch: *Gefahr*) bezeichnet die Wahrscheinlichkeit, dass innerhalb eines definierten Zeitraums ein bestimmtes Ereignis (beispielsweise: Tod, Herzinfarkt) eintritt. Die *Hazard Ratio* (HR) ist der dimensionslose numerische Quotient aus den Hazard Raten zweier verglichener Untersuchungsgruppen [5]:

- $HR = 1$ bedeutet, dass kein Unterschied zwischen zwei Gruppen hinsichtlich eines bestimmten Ereignisses besteht
- $HR > 1$ bedeutet, dass für die Interventionsgruppe innerhalb eines definierten Zeitraumes ein größeres Risiko besteht, dass ein bestimmtes Ereignis eintritt als für die Kontroll- beziehungsweise Vergleichsgruppe
- $HR < 1$ bedeutet, dass das Risiko für ein bestimmtes Ereignis innerhalb eines definierten Zeitraumes für die Interventionsgruppe kleiner ist als in der Kontroll- beziehungsweise Vergleichsgruppe

In Überlebenszeitanalysen bezeichnet die Hazard die momentane Sterberate in einer Patientengruppe. Die Hazard Ratio gibt an, um wie viel die Sterberate in der Interventionsgruppe höher oder niedriger ist als in der verglichenen Gruppe [85].

Die Berechnung der Hazard Ratios mit ihren jeweiligen 95 %-Konfidenzintervallen der für diese Arbeit Neuberechneten Metaanalysen I und II wurde nicht aus den Originaldaten der einzelnen Primärstudien wiederholt, sondern aus der Metaanalyse von *Kidane et al.* entnommen (Abbildung 1) [34].

Konfidenzintervall

Das Konfidenzintervall ist ein sogenanntes Vertrauensintervall, das den Bereich angibt, der bei unendlicher Wiederholung eines Experiments oder einer Untersuchung die wahre Lage eines Parameters miteinschließt. Das häufig verwendete Konfidenzniveau von 95 % (95%-CI) gibt an, dass bei identischer Wiederholung des Experiments das Intervall in näherungsweise 95 % der Fälle den wahren Parameter beinhalten wird.

Die Weite des Vertrauensbereichs hängt von Stichprobengröße und Standardabweichung der untersuchten Gruppen ab. Eine große Stichprobe führt zu „mehr Vertrauen“ und demnach zu einem engeren Konfidenzintervall. Bei kleinen Studiengruppen und großer Streuung der Werte wird die Aussage unsicherer und das Konfidenzintervall breiter [22]. Es gibt somit die Präzision der Lageschätzung eines Parameters an.

Konfidenzintervalle geben auch Aufschluss über die Richtung und die statistische Signifikanz des zu untersuchenden Effekts [22]:

- enthält der Vertrauensbereich die Zahl 1 nicht, so kann von einem *statistisch signifikanten* Ergebnis ausgegangen werden
- enthält der Vertrauensbereich die Zahl 1, so kann von einem *statistisch nicht signifikanten* Ergebnis ausgegangen werden
- liegt das CI vollständig unterhalb der Zahl 1 kann von einem *protektiven Effekt* ausgegangen werden
- liegt das CI vollständig oberhalb der Zahl 1, ist eine *Risikoerhöhung* zu erwarten

Im zweiten Schritt wird eine mögliche Heterogenität identifiziert und aus den Effektgrößen der einzelnen Studien eine gemeinsame Effektgröße (sogenannter *Meta-Effektschätzer* oder *Gesamteffekt*) mit dem zugehörigen Präzisionsmaß in Form eines 95 %-Konfidenzintervalles geschätzt.

Heterogenität und statistische Modelle

Obwohl alle in eine Metaanalyse eingeschlossenen Studien eine gemeinsame Fragestellung beantworten, sind oft Unterschiede in den Studienergebnissen festzustellen, die auf methodische, klinische oder statistische Faktoren wie verschiedene Studiendesigns, unterschiedliche Studienpopulationen oder Studiendurchführungen zurückgeführt werden können [18]. Aus diesem Grund ist es erforderlich, eine mögliche Heterogenität zwischen den einzelnen eingeschlossenen Studien zu identifizieren und zu quantifizieren, um eine objektivere Interpretation der Meta-Effektschätzung zu garantieren.

In dieser Arbeit wurde mithilfe der Statistiksoftware R, Version 3.2.0., mit dem Paket *Meta*, das Maß I^2 nach Higgins/Thompson berechnet, das auf einen Wertebereich zwischen 0 und 100 % normiert ist [29]. I^2 kann als derjenige Anteil der gesamten Variabilität der Effektschätzungen definiert werden, der eher auf Inhomogenität als auf reinen Zufall zurückzuführen ist [18].

Im Cochrane Handbuch für *Systematic Reviews of Interventions* wird ein Leitfaden zur Interpretation der I^2 -Statistik vorgestellt [18]:

- 0 % bis 40 %: wahrscheinlich keine Heterogenität;
- 30 % bis 60 %: könnte eine moderate Heterogenität darstellen
- 50 % bis 90 %: könnte erhebliche Heterogenität darstellen
- 75 % bis 100 %: erhebliche Heterogenität

Auf Grundlage dieser Heterogenitätsanalyse wurde eine Entscheidung bezüglich des passenden statistischen Modells zur weiteren Bestimmung des Meta-Effektschätzers getroffen: Die Daten können mittels eines Festeffektmodells (*Fixed-Effects Modell*) oder Zufallseffektmodells (*Random-Effects Modell*) ausgewertet werden.

Da keine empfohlene Vorgehensweise für die Auswahl eines der zwei Modelle existiert und das Cochrane Institut empfiehlt, diese nicht allein auf Grundlage von statistischen Tests zu treffen [18], wurde der Meta-Schätzwert der Neuberechneten Metaanalysen I und II jeweils mit beiden Modellen bestimmt.

Zunächst wurden die Metaanalysen unter Annahme eines Festeffektmodells (*Fixed-Effects Modell*) durchgeführt, das eine mögliche Heterogenität zwischen den einzelnen eingeschlossenen Studien ignoriert und eine hohe Vergleichbarkeit annimmt [36]. Eine Abweichung zwischen den Effektschätzungen der einzelnen Studien hat keine Auswirkungen auf das Konfidenzintervall [18]. Bei der Berechnung des 95 %-Konfidenzintervalls wird somit lediglich die studieninterne Variabilität (*Intra-Studien-Varianz*) hinsichtlich des kombinierten Effekts der Studien berücksichtigt und auf dessen Grundlage den jeweiligen Studien eine bestimmte Gewichtung zugeschrieben [11].

Um nicht zu schmale Konfidenzintervalle zu erhalten, die eine falsch-positive Zuverlässigkeit suggerieren könnten [11], wurde die Inter-Studien-Varianz τ^2 (τ^2 , *Heterogenität zwischen den Studien*) nach dem Schätzer von *DerSimonian und Laird* berechnet [19] und die Metaanalysen zusätzlich unter Annahme eines Zufallseffektmodells (*Random-Effects Modell*) durchgeführt. Dieses Modell berücksichtigt in der Berechnung des 95 %-Konfidenzintervalles und folglich der Gewichtung der einzelnen Studien sowohl die Variabilität innerhalb der Studien (*Intra-Studien-Varianz*) als auch zwischen den Studien (*Inter-Studien-Varianz*). Der Meta-Schätzwert wird somit unter Korrektur der Heterogenität zwischen den einzelnen Studien berechnet [11].

Forest Plot

Alle Ergebnisse und Kenngrößen der Metaanalysen I und II wurden in einem Forest Plot graphisch und numerisch dargestellt (Abbildung 4 und Abbildung 5).

Ganz links stehen die eingeschlossenen Studien mit Namen und Jahr sowie Patientenanzahl für die Interventionsgruppe N(T) und Kontrollgruppe N(C). Die Hazard Ratio wird als Vertreter der Effektgröße als gefülltes Quadrat und das 95 %-Konfidenzintervall als Verlängerungen ausgehend von diesem Quadrat nach rechts und links dargestellt. Die unterschiedlichen Größen der gefüllten Quadrate repräsentieren die unterschiedlichen Gewichtungen der Einzelstudien, die ihnen in der Metaanalyse zugeordnet werden. Das jeweilige numerische Ausmaß der Gewichtung der einzelnen Studien unter Berücksichtigung eines Festeffektmodells $W(\text{fixed})$ und eines Zufalleffektmodells $W(\text{random})$ wird prozentual ganz rechts abgebildet. Des Weiteren wird rechts die HR als numerische Effektgröße mit dem jeweiligen 95 %-CI dargestellt. Die Meta-Effektgrößen der zwei Metaanalysen wird durch eine Raute (sogenannter *diamond*) ausgedrückt, deren Lage der Gesamteffektgröße und deren Breite der Präzision der Schätzung entspricht. Eine Angabe erfolgt sowohl unter Annahme eines Festeffektmodells als auch eines Zufalleffektmodells. Die Heterogenität wird entsprechend ihrer Kennwerte I^2 , τ^2 und p dokumentiert.

Für die Berechnung der Metaanalysen wurde die Statistiksoftware R, Version 3.2.0, mit dem *Paket Meta* verwendet.

Abbildung 3 veranschaulicht die Methodik der Validitätsbewertung:

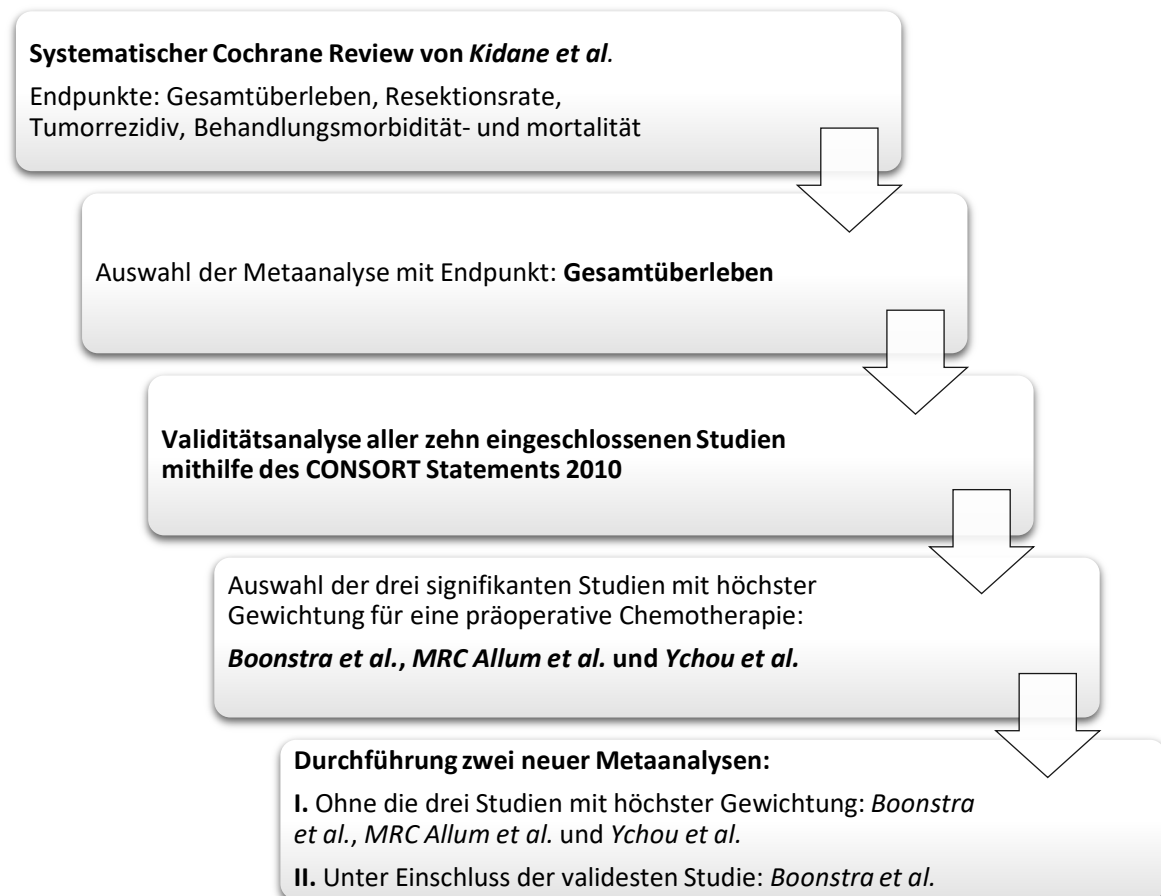


Abbildung 3: Die fünf Schritte der Validitätsbewertung der Metaanalyse Gesamtüberleben des Cochrane Reviews von Kidane et al. Die Metaanalyse mit Endpunkt Gesamtüberleben wurde ausgewählt und die Validität aller zehn Einzelstudien mithilfe des CONSORT Statements 2010 beurteilt. Von den signifikanten Studien wurden die drei mit höchster Gewichtung ausgewählt und zwei neue Metaanalysen durchgeführt.

Quelle: Bei Abbildung 3 handelt es sich um eine Eigene Darstellung. Universitätsklinikum Ulm 2021.

3 Ergebnisse

3.1 Validitätsbewertung der signifikanten Studien

Die Ergebnisse der Auswertung der vier signifikanten Studien *Boonstra et al.*, *Law et al.*, *MRC Allum et al.* und *Ychou et al.* mithilfe der CONSORT Checkliste 2010 werden in Tabelle 3 dargestellt.

Tabelle 3: Validitätsbewertung der signifikanten Studien mithilfe der CONSORT Checkliste 2010

Tabelle 3 fasst alle 37 Validitätskriterien der CONSORT Checkliste 2010 zusammen und präsentiert die Ergebnisse der Validitätsbewertung der vier signifikanten Studien *Boonstra et al.*, *Law et al.*, *MRC Allum et al.* und *Ychou et al.* In der letzten Zeile wird die Anzahl der nicht erfüllten Kriterien in Absolutzahl und Prozentzahl von den Gesamtkriterien minus den Kriterien *N.anwendbar* angegeben.

Abkürzungen: *N.anwendbar*: Nicht anwendbar

Quelle: Modifiziert nach Manzini G, Klotz U, Henne-Bruns D, Kremer M: *Validity of studies suggesting preoperative chemotherapy for resectable thoracic esophageal cancer: A critical appraisal of randomized trials.* World Journal of Gastrointestinal Oncology 12: S. 118 (2020) [48], DOI: 10.4251/wjgo.v12.i1.113 und wurde unter CC BY-NC 4.0 (<http://creativecommons.org/licenses/by-nc/4.0/>) publiziert.

Thema	Nummer	Boonstra 2011	Law 1997	MRC Allum 2009	Ychou 2011
<i>Titel</i>	1a	Ja	Ja	Ja	Nein
<i>Abstract</i>	1b	Nein	Nein	Nein	Nein
Einleitung					
<i>Hintergrund</i>	2a	Ja	Ja	Ja	Ja
<i>Fragestellung</i>	2b	Ja	Ja	Ja	Ja
Methoden					
<i>Studiendesign</i>	3a	Ja	Ja	Ja	Ja
<i>Methodenänderungen nach Studienbeginn</i>	3b	N.anwendbar	N.anwendbar	N.anwendbar	Ja
<i>Eignungskriterien</i>	4a	Ja	Ja	Ja	Ja
<i>Ort der Durchführung</i>	4b	Ja	Nein	Nein	Nein
<i>Intervention</i>	5	Ja	Ja	Nein	Ja
<i>Endpunkte</i>	6a	Ja	Nein	Ja	Ja
<i>Endpunktänderungen nach Studienbeginn</i>	6b	N.anwendbar	N.anwendbar	N.anwendbar	N.anwendbar
<i>Fallzahlberechnung</i>	7a	Ja	Ja	Nein	Ja
<i>Zwischenanalysen und Abbruchkriterien</i>	7b	N.anwendbar	Nein	N.anwendbar	Ja
Randomisierung					
<i>Zufallsgenerierung</i>	8a	Nein	Nein	Nein	Ja
<i>Art der Randomisierung</i>	8b	Nein	Nein	Nein	Nein
<i>Geheimhaltung</i>	9	Nein	Nein	Nein	Nein
<i>Durchführung</i>	10	Nein	Nein	Nein	Nein
<i>Verblindung</i>	11a	Nein	Nein	Nein	Nein
<i>Beschreibung der Ähnlichkeit der Interventionen</i>	11b	Nein	Nein	Nein	Nein
<i>Statistische Methoden</i>	12a	Ja	Ja	Ja	Ja
<i>Methoden zusätzlicher Analysen</i>	12b	Ja	N.anwendbar	Ja	N.anwendbar

Thema	Nummer	Boonstra 2011	Law 1997	MRC Allum 2009	Ychou 2011
Ergebnisse					
Teilnehmerflussdiagramm: Studienteilnehmer	13a	Ja	Nein	Ja	Ja
Studienausscheider	13b	Ja	Ja	Ja	Ja
Zeitraum der Rekrutierung und Nachbeobachtung	14a	Ja	Ja	Ja	Ja
Grund für Studienende	14b	N.anwendbar	N.anwendbar	N.anwendbar	Ja
Patientencharakteristika	15	Ja	Ja	Ja	Ja
Anzahl ausgewertete Pro- banden	16	Ja	Ja	Ja	Ja
Ergebnisse mit Effekt- größe	17a	Ja	Ja	Ja	Ja
bei binären Endpunkten: absolute und relative Ef- fektgrößen	17b	Ja	Ja	Ja	Ja
Zusätzliche Analysen	18	Ja	N.anwendbar	Ja	N.anwendbar
Schäden	19	Ja	Ja	Nein	Ja
Diskussion					
Limitierungen	20	Ja	Ja	Ja	Ja
Generalisierbarkeit	21	Nein	Nein	Nein	Nein
Interpretation	22	Ja	Ja	Ja	Ja
Andere Information					
Registrierung	23	Nein	Nein	Nein	Nein
Protokoll	24	Nein	Nein	Nein	Nein
Finanzierung	25	Ja	Nein	Ja	Nein
Anzahl nicht erfüllter CONSORT Kriterien (prozentualer Anteil)		10 (30,3 %)	15 (46,8 %)	14 (42,4 %)	12 (35,2 %)

In Tabelle 4 werden die wesentlichen Merkmale der drei Primärstudien mit der höchsten Gewichtung, *Boonstra et al.*, *MRC Allum et al.* und *Ychou et al.*, zusammengefasst.

Tabelle 4: Zusammenfassung wesentlicher Merkmale der Studien Boonstra et al., MRC Allum et al. und Ychou et al.

Abkürzungen: AEG: Adenokarzinom des gastroösophagealen Überganges, CI: Konfidenzintervall, CTx: Chemotherapie, HR: Hazard Ratio, iv.: intravenös, OP: Operation, p: p-Wert, po.: per os, präop.: präoperativ, T N M: Tumor Lymphknoten Metastasen (gemäß TNM Klassifikation), vs.: versus, WHO: Weltgesundheitsorganisation, 5-FU: 5-Fluoruracil

Quelle: Modifiziert nach Manzini G, Klotz U, Henne-Bruns D, Kremer M: *Validity of studies suggesting preoperative chemotherapy for resectable thoracic esophageal cancer: A critical appraisal of randomized trials.* World Journal of Gastrointestinal Oncology 12: S. 117 (2020) [48], DOI: 10.4251/wjgo.v12.i1.113 und wurde unter CC BY-NC 4.0 (<http://creativecommons.org/licenses/by-nc/4.0/>) publiziert.

Thema	Boonstra (2011)	MRC Allum (2009)	Ychou (2011)
Patientenanzahl (Intervention vs. Kontrolle)	85 vs. 84	400 vs. 402	113 vs. 111
Eignungskriterien	100 % Plattenepithelkarzinom des intrathorakalen Ösophagus (oberes, mittleres und unteres Drittel), T1-3, jedes N, M0 (M1a wenn distaler Ösophagus betroffen UND vermutlicher Lymphknotenbefall des Truncus coeliacus, <80 Jahre, Karnofsky-Index >70.	Plattenepithelkarzinom, Adenokarzinom, undifferenziertes Karzinom Oberes, mittleres und unteres Drittel des Ösophagus mit Kardia des Magens	Adenokarzinom des unteren Ösophagusdrittels, gastroösophagealen Übergangs oder Magens (AEG-Tumor), 18-75 Jahre, WHO Performance Status 0 oder 1, adäquate Nierenfunktion (Kreatinin < 120 µmol/L) und adäquate hämatologische Funktionen
Interventionsgruppe	Präop. CTx: Cisplatin, Etoposid iv. & po. + OP	Präop. CTx: Cisplatin, 5-FU + präop. Radiatio + OP	Präop. CTx: 5-FU, Cisplatin + OP
Kontrollgruppe	OP	Präop. Radiatio + OP	OP
Ergebnis (Intervention vs. Kontrolle)	Medianes Gesamtüberleben 16 Monate vs. 12 Monate HR 0.71; (95 %-CI 0.51 bis 0.98, p=0.03)	Medianes Gesamtüberleben war signifikant länger in der Interventionsgruppe HR 0.84; (95 %-CI 0.72 bis 0.98, p=0.03)	Medianes Gesamtüberleben war signifikant länger in der Interventionsgruppe HR 0,65; (95 %-CI 0.50 bis 0.95, p=0.02) 5 Jahresüberleben war 38 % (95 %-CI 29 % – 47 %) in der Interventionsgruppe vs. 24 % (95 %-CI 26 % – 44 %) in der Kontrollgruppe.
Gewichtung	24.1 %	20.5 %	24.5 %

Im Folgenden werden die Ergebnisse der Validitätsbewertung der vier signifikanten Studien thematisch näher erläutert:

3.1.1 Ychou et al. 2011

In der Studie von *Ychou et al.* werden gemäß der CONSORT Checkliste 12 von 34 Validitätskriterien (35,2 %) nicht erfüllt. Drei Kriterien sind nicht anwendbar: *Endpunktänderungen nach Studienbeginn*, *Methoden zusätzlicher Analysen* und *zusätzliche Analysen*

Randomisierung und Verblindung

Die Randomisierung zur Interventions- oder Kontrollgruppe erfolgte mithilfe eines Telefonanrufs über ein zentrales Randomisierungssystem. Die Zuordnung wurde mithilfe der Minimierungsmethode nach Zentrum, WHO Performance Status und Tumorlokalisation (Ösophagus-, AEG-*, und Magentumore) geschichtet. Durch die Verwendung einer solchen dynamischen Randomisierungsmethode konnte eine adäquate Geheimhaltung der Randomisierungsliste (sogenanntes *allocation concealment*) nicht mehr gewährleistet werden.

Die fehlende Implementierung einer präoperativen Behandlung von Patienten in der Kontrollgruppe machte eine Verblindung von Studienteilnehmer, Studienärzten, Bewerter und somit eine Beschreibung der Ähnlichkeit der Interventionen unmöglich.

Methoden

Im Jahr 1998 wurde das ursprüngliche Studiendesign mit seinen Eignungskriterien verändert und es wurden nachträglich 55 Patienten (25 % aller Patienten, insgesamt 28 in der Interventions- und 27 in der Kontrollgruppe) mit Adenokarzinom des Magens in die Studie eingeschlossen.

Gemäß den statistischen Berechnungen der Studienautoren hätte die Studie 250 Patienten (mit 178 auftretenden Todesfälle) benötigt, um die erforderliche Teststärke (sogenannte *power*) zu erreichen. Da die Studie aufgrund erschwerter Patientenrekrutierung 2003 verfrüht geschlossen wurde, konnten schlussendlich

* Adenokarzinom des Gastroösophagealen Überganges: von Barrett-Karzinomen (Typ I) über Kardial-Karzinomen (Typ II) zu Subkardial-Karzinomen, eigentlich Magenkarzinomen (Typ III), reichend

lediglich 224 Patienten (mit 156 Todesfällen im Verlauf) in die Studie eingeschlossen werden. Dies stellt die Gewährleistung einer ausreichenden Teststärke infrage.

Der genaue Ort und das Umfeld der Studie beziehungsweise der Interventionen wurden in der Studie nicht näher angegeben.

Zusätzliche Therapiemaßnahmen

Bei guter Toleranz und fehlendem Hinweis auf Tumorprogress nach präoperativer Chemotherapie erhielten 50 % der Patienten der Interventionsgruppe bis zu 4 Zyklen postoperativer Chemotherapie mit 5-FU und Cisplatin.

3.1.2 MRC Allum et al. 2009

Die Studie von *MRC Allum et al.* beschreibt die Langzeitergebnisse einer zuvor veröffentlichten Studie [49] derselben Gruppe aus dem Jahr 2002. Wenn benötigte Informationen in der neueren Studie von *MRC Allum et al.* fehlten, wurde überprüft, ob diese in der ersten Veröffentlichung verfügbar waren [49]. Somit werden 14 von 33 Validitätskriterien (42,4 %) nicht erfüllt. Vier Kriterien sind nicht anwendbar: *Methodenänderungen nach Studienbeginn, Endpunktänderungen nach Studienbeginn, Zwischenanalysen, Abbruchkriterien* und *Grund für Studienende*.

Randomisierung und Verblindung

In dieser Studie wurde die Methode und Art der Generierung einer Zufallsliste nicht näher beschrieben. Auch hier wurde die Minimierung als alternative stratifizierende Randomisation verwendet und die Patienten nach operierenden Chirurgen, Tumorkomlokalisierung (oberes, mittleres, unteres Drittel des Ösophagus oder Kardia), WHO Performance Status und Histopathologie (undifferenziertes, Platten-, oder Adenokarzinom) geschichtet. Demzufolge konnte bei einer solchen Methodik eine adäquate Geheimhaltung der Randomisierungsliste nicht sichergestellt werden.

Eine Verblindung wurde nicht durchgeführt, da die Kontrollgruppe keinerlei präoperative Therapie mit einem Placebo erhielt.

Methoden

Die konkrete Durchführung der Therapie kann nicht nachvollzogen werden, da die genaue Operationsart nicht beschrieben wurde. Je nach Tumorlokalisation und individueller Expertise wurde seitens der Operateure das chirurgische Verfahren gewählt.

Eine Fallzahlplanung oder rationale Begründung der Stichprobengröße fehlten.

Der genaue Ort und das Umfeld der Studie beziehungsweise der Interventionen wurden auch in dieser Studie nicht näher angegeben.

Zusätzliche Therapiemaßnahmen

9 % (n=36) der Patienten sowohl der Interventions- als auch Kontrollgruppe erhielten eine präoperative Bestrahlung mit 25 Gy fraktioniert über eine Woche, 32,5 Gy fraktioniert über zwei Wochen oder einer äquivalenten Dosis.

Ergebnisse

Die Studie beschrieb im Gegensatz zu den anderen zwei zuvor genannten Studien keinerlei Toxizität oder Nebenwirkungen der Interventionen.

In Bezug auf das Patientenalter und den Grad der Dysphagie unterschieden sich die beiden Gruppen mit einer gewissen Wahrscheinlichkeit einer unausgewogenen Risikoverteilung zugunsten der Interventionsgruppe.

Weitere Informationen

In der ersten Veröffentlichung vom Jahr 2002 legten die Studiensponsoren das Komitee fest, das schlussendlich die Daten interpretierte und die Publikation verfasste [49].

3.1.3 Boonstra et al. 2011

In der Studie von *Boonstra et al.* zeigen sich in 10 von 33 Kriterien eine mangelnde Validität (30,3 %). Vier Kriterien sind nicht anwendbar: *Methodenänderungen nach Studienbeginn, Endpunktänderungen nach Studienbeginn, Zwischenanalysen, Abbruchkriterien* und *Grund für Studienende*.

Randomisierung und Verblindung

Der Studienbericht gab keinen Aufschluss darüber, wie die Randomisierung durchgeführt wurde, sondern lediglich, dass sie am medizinischen Zentrum der Erasmus Universität in Rotterdam umgesetzt worden sei. Aus diesem Grund muss eine mangelnde Geheimhaltung der Randomisierungsliste angenommen werden. Die Patienten wurden anhand von Alter, Geschlecht, Gewichtsverlust in den letzten 4 Monaten und Tumorgröße stratifiziert.

Wie in den bereits vorab erwähnten Studien wurde auch in dieser Studie keine Verblindung der Teilnehmer realisiert.

Ergebnisse

Aufgrund fehlender oder unklarer Information bezüglich Studienabbrecher (sogenannte *Drop-outs*) wird aus der Studienbeschreibung nicht ersichtlich, ob eine Intention-to-treat-Analyse durchgeführt wurde oder nicht.

3.1.4 Law et al. 1997

In der Studie von *Law et al.* werden 15 von 32 Validitätskriterien nicht erfüllt (46,8 %). Fünf Kriterien sind nicht anwendbar: *Methodenänderungen nach Studienbeginn, Endpunktänderungen nach Studienbeginn, Methoden zusätzlicher Analysen, Grund für Studienende* und *Zusätzliche Analysen*.

Randomisierung und Verblindung

Im Studienreport wurde weder die Methode der Zufallsgenerierung noch die Art der Randomisierung dargelegt.

Die fehlende Implementierung einer präoperativen Behandlung der Patienten in der Kontrollgruppe machte eine Verblindung und somit eine Beschreibung der Ähnlichkeit der Interventionen unmöglich.

Methoden

Für den Leser dieses Studienberichts ist nicht klar ersichtlich, wie die genauen primären und sekundären Endpunkte definiert wurden. Es wurde lediglich aufgezeigt, dass die Endpunkte tumor- und therapiebezogene Todesfälle sind.

Es fehlen zudem Angaben über den genauen Ort und dem Umfeld der Studie beziehungsweise der Interventionen.

Ergebnisse

Ein vom CONSORT Statement gefordertes Teilnehmerflussdiagramm oder dessen Äquivalent dazu wurde nicht veröffentlicht. Studienteilnehmer und Studienausscheider sind somit nicht schlüssig nachvollziehbar.

3.2 Validitätsbewertung der nicht-signifikanten Studien

Die Ergebnisse der sechs Studien (*Ancona et al.* (2001), *Kelsen et al.* (1998), *Maipang et al.* (1994), *Nygaard et al.* (1992), *Roth et al.* (1988) und *Schlag et al.* (1992)), die keinen signifikanten Unterschied im Gesamtüberleben von Patienten mit präoperativer Chemotherapie feststellen konnten, werden in Tabelle 5 vorgestellt.

In den ältesten dieser nicht-signifikanten Studien von *Maipang et al.*, *Nygaard et al.*, *Roth et al.* und *Schlag et al.* werden über 50 % der CONSORT Kriterien nicht erfüllt. Die neueren Studien von *Ancona et al.* und *Kelsen et al.* erfüllen über 40 % der Items der CONSORT Checkliste 2010 nicht.

Im Folgenden werden die Ergebnisse der Validitätsanalyse der nicht-signifikanten Studien zusammenfassend näher erläutert.

Randomisierung und Verblindung

Die Art der Generierung einer Zufallsliste wurde in keiner der Studien beschrieben. Lediglich die Studie von *Ancona et al.* beschrieb die Art der Randomisierung, die als Blockrandomisation durchgeführt wurde.

Eine adäquate Geheimhaltung der Randomisierungsliste konnte von keiner Studie sichergestellt werden.

Allen Studien fehlte die Implementierung einer präoperativen Behandlung von Patienten in der Kontrollgruppe und machte eine Verblindung von Studienteilnehmern,

Studienärzten, Bewertern und somit eine Beschreibung der Ähnlichkeit der durchgeführten Interventionen unmöglich.

Methoden

Die primären und sekundären Endpunkte wurden nur in den neueren Studien von *Ancona et al.* und *Kelsen et al.* vollständig definiert und fehlten in den übrigen vier Studien völlig.

Nur zwei Studien, *Kelsen et al.* und *Schlag et al.*, gaben die Berechnungsgrundlage der Stichprobengrößen an.

Ergebnisse

Das vom CONSORT Statement 2010 empfohlene Flussdiagramm, über nachvollziehbare Ein- und Ausschlüsse von Teilnehmern, wurde in keiner der Studien dargestellt.

In der Studie von *Roth et al.* fehlte eine Schilderung von demographischen und klinischer Patientencharakteristika. Alle übrigen Studien beschrieben in ausreichendem Maße sämtliche Patientencharakteristika.

In der Studie von *Schlag et al.* wurde die Dauer des Nachbeobachtungsintervalls nicht angegeben, sondern lediglich die durchgeführten Untersuchungen.

Zusätzliche Therapiemaßnahmen

In der Studie von *Ancona et al.* erhielten Patienten sowohl der Interventionsgruppe als auch Kontrollgruppe zusätzlich postoperative Chemotherapie und Radiotherapie bei Vorliegen eines postoperativ inkompletten Resektionsstatus (R1 und R2).

In der Studie von *Kelsen et al.* erhielten Patienten in der Interventionsgruppe postoperative Chemotherapie, wenn sie darauf angesprochen haben, und Bestrahlung bei tumorpositiven Resektionsrändern.

In der Studie von *Nygaard et al.* wurden vier Interventionsgruppen gebildet und die Patienten zugeteilt: 1. *Operation*, 2. *Präoperative Chemotherapie + Operation*, 3. *Präoperative Radiatio + Operation*, 4. *Präoperative Chemotherapie + Radiatio + Operation*.

Tabelle 5: Validitätsbewertung der nicht- signifikanten Studien mithilfe der CONSORT Checkliste 2010

Die Tabelle fasst alle 37 Validitätskriterien der CONSORT Checkliste 2010 zusammen und präsentiert die Ergebnisse der Validitätsbewertung der sechs nicht-signifikanten Studien: *Ancona et al.*, *Kelsen et al.*, *Maipang et al.*, *Nygaard et al.*, *Roth et al.* und *Schlag et al.* In der letzten Zeile werden die Anzahl der nicht erfüllten Kriterien in Absolutzahl und Prozentzahl von den Gesamtkriterien minus den Kriterien *N.anwendbar* angegeben.

Abkürzungen: *N.anwendbar*: Nicht anwendbar

Quelle: Bei Tabelle 5 handelt es sich um eine Eigene Darstellung. Universitätsklinikum Ulm 2020.

Thema	Nummer	Ancona 2001	Kelsen 1998	Maipang 1994	Nygaard 1992	Roth 1988	Schlag 1992
<i>Titel</i>	<i>1a</i>	Ja	Nein	Nein	Ja	Ja	Ja
<i>Abstract</i>	<i>1b</i>	Nein	Nein	Nein	Nein	Nein	Nein
Einleitung							
<i>Hintergrund</i>	<i>2a</i>	Ja	Ja	Ja	Ja	Ja	Ja
<i>Fragestellung</i>	<i>2b</i>	Ja	Ja	Ja	Ja	Ja	Ja
Methoden							
<i>Studiendesign</i>	<i>3a</i>	Ja	Ja	Ja	Ja	Ja	Ja
<i>Methodenänderungen nach Studienbeginn</i>	<i>3b</i>	N.anwendbar	Ja	N.anwendbar	N.anwendbar	N.anwendbar	N.anwendbar
<i>Eignungskriterien</i>	<i>4a</i>	Ja	Ja	Ja	Ja	Ja	Ja
<i>Ort der Durchführung</i>	<i>4b</i>	Nein	Nein	Ja	Nein	Nein	Nein
<i>Intervention</i>	<i>5</i>	Ja	Ja	Ja	Ja	Ja	Ja
<i>Endpunkte</i>	<i>6a</i>	Ja	Ja	Nein	Nein	Nein	Nein
<i>Endpunktänderungen nach Studienbeginn</i>	<i>6b</i>	N.anwendbar	N.anwendbar	N.anwendbar	N.anwendbar	N.anwendbar	N.anwendbar
<i>Fallzahlbestimmung</i>	<i>7a</i>	Nein	Ja	Nein	Nein	Nein	Ja
<i>Zwischenanalysen und Abbruchkriterien</i>	<i>7b</i>	N.anwendbar	Ja	N.anwendbar	N.anwendbar	N.anwendbar	Nein
Randomisierung							

Thema	Nummer	Ancona 2001	Kelsen 1998	Maipang 1994	Nygaard 1992	Roth 1988	Schlag 1992
<i>Zufallsgenerierung</i>	<i>8a</i>	Nein	Nein	Nein	Nein	Nein	Nein
<i>Art der Randomisierung</i>	<i>8b</i>	Ja	Nein	Nein	Nein	Nein	Nein
<i>Geheimhaltung</i>	<i>9</i>	Nein	Nein	Nein	Nein	Nein	Nein
<i>Durchführung</i>	<i>10</i>	Nein	Nein	Nein	Nein	Nein	Nein
<i>Verblindung</i>	<i>11a</i>	Nein	Nein	Nein	Nein	Nein	Nein
<i>Beschreibung der Ähnlichkeit der Interventionen</i>	<i>11b</i>	Nein	Nein	Nein	Nein	Nein	Nein
<i>Statistische Methoden</i>	<i>12a</i>	Ja	Ja	Ja	Ja	Ja	Ja
<i>Methoden zusätzlicher Analysen</i>	<i>12b</i>	Ja	N.anwendbar	N.anwendbar	N.anwendbar	N.anwendbar	N.anwendbar
Ergebnisse							
<i>Teilnehmerflussdiagramm</i>	<i>13a</i>	Nein	Nein	Nein	Nein	Nein	Nein
<i>Studienausscheider</i>	<i>13b</i>	Ja	Ja	Ja	Ja	Ja	Ja
<i>Zeitraum der Rekrutierung und Nachbeobachtung</i>	<i>14a</i>	Ja	Ja	Ja	Ja	Ja	Nein
<i>Grund für Studienende</i>	<i>14b</i>	Ja	Nein	Nein	Nein	Nein	Nein
<i>Patientencharakteristika</i>	<i>15</i>	Ja	Ja	Ja	Ja	Nein	Ja
<i>Anzahl ausgewertete Probanden</i>	<i>16</i>	Ja	Ja	Ja	Ja	Ja	Ja
<i>Ergebnisse mit Effektgröße bei binären Endpunkten: absolute und relative Effektgrößen</i>	<i>17a</i>	Nein	Ja	Nein	Nein	Nein	Nein
	<i>17b</i>	Ja	Ja	Ja	Ja	Ja	Ja
<i>Zusätzliche Analysen</i>	<i>18</i>	N.anwendbar	N.anwendbar	N.anwendbar	N.anwendbar	N.anwendbar	N.anwendbar
<i>Schaden</i>	<i>19</i>	Ja	Ja	Ja	Ja	Ja	Ja

<i>Thema</i>	<i>Nummer</i>	<i>Ancona 2001</i>	<i>Kelsen 1998</i>	<i>Maipang 1994</i>	<i>Nygaard 1992</i>	<i>Roth 1988</i>	<i>Schlag 1992</i>
<i>Diskussion</i>							
<i>Limitierungen</i>	20	Ja	Ja	Ja	Ja	Ja	Ja
<i>Generalisierbarkeit</i>	21	Nein	Nein	Nein	Nein	Nein	Nein
<i>Interpretation</i>	22	Ja	Ja	Ja	Ja	Ja	Ja
<i>Andere Information</i>							
<i>Registrierung</i>	23	Nein	Nein	Nein	Nein	Nein	Nein
<i>Protokoll</i>	24	Nein	Nein	Nein	Nein	Nein	Nein
<i>Finanzierung</i>	25	Nein	Ja	Ja	Nein	Nein	Nein
<i>Anzahl nicht erfüllter CONSORT Kriterien (pro- zentualer Anteil)</i>		14 (42,4 %)	14 (41,1 %)	16 (50 %)	17 (53,1 %)	18 (56,2 %)	18 (54,5 %)

3.3 Ergebnisse der neuen Metaanalysen

Aufgrund des im vorangegangenen Ergebnisteil aufgezeigten Validitätsdefizites der Studien von *Boonstra et al.*, *MRC Allum et al.* und *Ychou et al.* wurden zwei neue Metaanalysen durchgeführt.

3.3.1 Metaanalyse I

Abbildung 4 veranschaulicht das Ergebnis der ersten Metaanalyse, in der die drei Studien mit der höchsten Gewichtung ausgeschlossen werden (*Boonstra et al.*, *MRC Allum et al.* und *Ychou et al.*). Insgesamt werden somit sieben Studien eingeschlossen. Eine eingeschlossene Studie (*Law et al.*) zeigte ein positives und statistisch signifikantes Ergebnis für eine präoperative Chemotherapie und wird mit 19,5 % unter Annahme eines Festeffektmodells und mit 21,1 % unter Annahme eines Zufallseffektmodells gewichtet. Die weiteren sechs Studien (*Ancona et al.*, *Kelsen et al.*, *Maipang et al.*, *Nygaard et al.*, *Roth et al.* und *Schlag et al.*) zeigten kein statistisch signifikantes Ergebnis.

Die Schätzung der Metaanalyse I ergibt eine HR von 0,94 (95 %-CI: 0,81 bis 1,09) unter der Annahme eines Festeffektmodells und eine HR von 0,92 (95 %-CI: 0,75 bis 1,13) unter der Annahme eines Zufallseffektmodells. Unerheblich welches Modell genutzt wird, kann der Gesamtüberlebensvorteil einer präoperativen Chemotherapie für Patienten mit resektablem thorakalen Ösophaguskarzinom durch die Metaanalyse I nicht mehr unterstützt werden. Die Schätzung der ursprünglichen Metaanalyse von *Kidane et al.* ergab eine HR von 0,88 (95 %-CI: 0,80 bis 0,96).

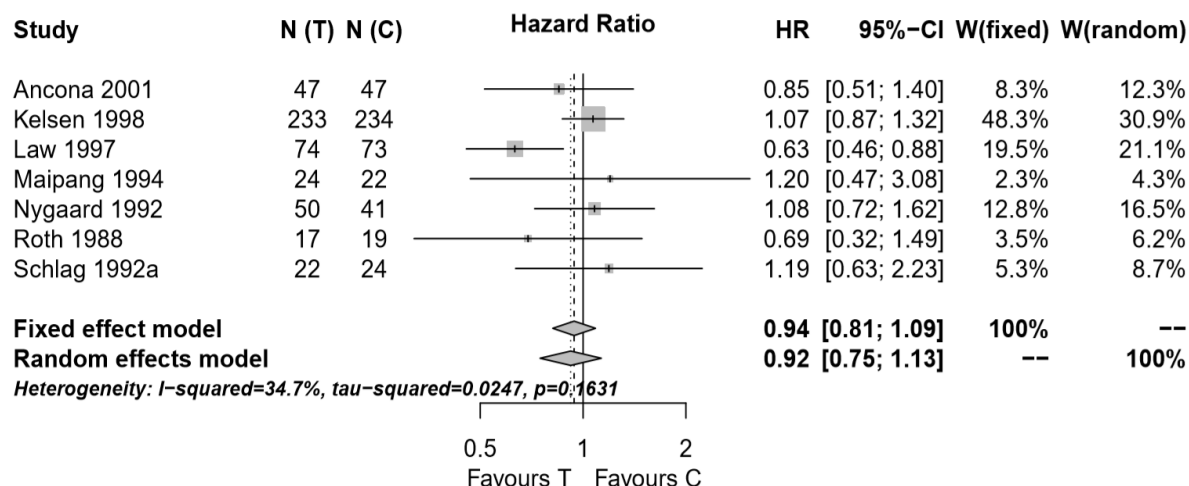


Abbildung 4: Metaanalyse I mit sieben Studien nach Ausschluss der Studien *Boonstra et al.*, *MRC Allum et al.* und *Ychou et al.*

Links stehen die sieben Studien mit der jeweiligen Patientenanzahl in der Interventions- und Kontrollgruppe. Mittig wird die Hazard Ratio als graues Quadrat und das 95 %-Konfidenzintervall als waagerechte Verlängerung graphisch in einem Forest-Plot dargestellt. Die grauen Rauten symbolisieren die Gesamteffektgröße sowohl unter Annahme eines Festeffektmodells als auch unter Annahme eines Zufallseffektmodells. Rechts werden die Gewichtung der Einzelstudien prozentual sowohl unter Annahme eines Festeffektmodells als auch unter Annahme eines Zufallseffektmodells und die Hazard Ratio mit dem 95 %-Konfidenzintervall numerisch angegeben. Die Heterogenität wird entsprechend ihrer Kennwerte I^2 , τ^2 und p angegeben.

Abkürzungen: C: Kontrollgruppe, HR: Hazard Ratio, $N(T)$: Anzahl der Patienten in der Versuchsgruppe, $N(C)$: Anzahl der Patienten in der Kontrollgruppe, p : P-Wert, T: Versuchsgruppe, $W(fixed)$: Gewicht, das der Studie unter Verwendung eines Festeffektmodells zugewiesen wurde, $W(random)$: Gewicht, das der Studie unter Verwendung eines Zufallseffektmodells zugewiesen wurde

Quelle: Abbildung 4 stammt aus Manzini G, Klotz U, Henne-Bruns D, Kremer M: *Validity of studies suggesting preoperative chemotherapy for resectable thoracic esophageal cancer: A critical appraisal of randomized trials*. *World Journal of Gastrointestinal Oncology* 12 (2020), S. 119 [48], DOI: 10.4251/wjgo.v12.i1.113 und wurde unter CC BY-NC 4.0 (<http://creativecommons.org/licenses/by-nc/4.0/>) publiziert.

3.3.2 Metaanalyse II

In der Studie von *Boonstra et al.* werden nur 10 von 33 Kriterien (30,3 %) der CONSORT Checkliste nicht erfüllt. Es wurde eine hinreichende Validität zum Einschluss in die Metaanalyse angenommen. Somit wurde eine zweite Metaanalyse, die acht Studien umfasst, unter Ausschluss der Studien von *MRC Allum et al.* und *Ychou et al.* durchgeführt (Abbildung 5).

Die Metaanalyse II beinhaltet nun zwei Studien mit positivem und signifikantem Ergebnis für eine präoperative Chemotherapie: die Studie von *Boonstra et al.* mit einer Gewichtung von 51,2 % unter Annahme eines Festeffektmodells und von 33,2 % unter Annahme eines Zufallseffektmodells und die Studie von *Law et al.* mit einer Gewichtung von 9,5 % unter Annahme eines Festeffektmodells und von 14,2 % unter Annahme eines Zufallseffektmodells.

Die Schätzung der Metaanalyse II ergibt eine HR von 0,90 (95 %-CI: 0,81 bis 1,00) unter Annahme eines Festeffektmodells und einer HR von 0,90 (95 %-CI: 0,78 bis 1,05) unter Annahme eines Zufallseffektmodells. Auch unter Einschluss der Studie mit der höchsten Validität aller analysierten Studien (*Boonstra et al.*) und unabhängig vom angenommenen Modell kann die neue Schätzung den Vorteil einer präoperativen Chemotherapie beim thorakalen resektablen Ösophaguskarzinom nicht bestätigen.

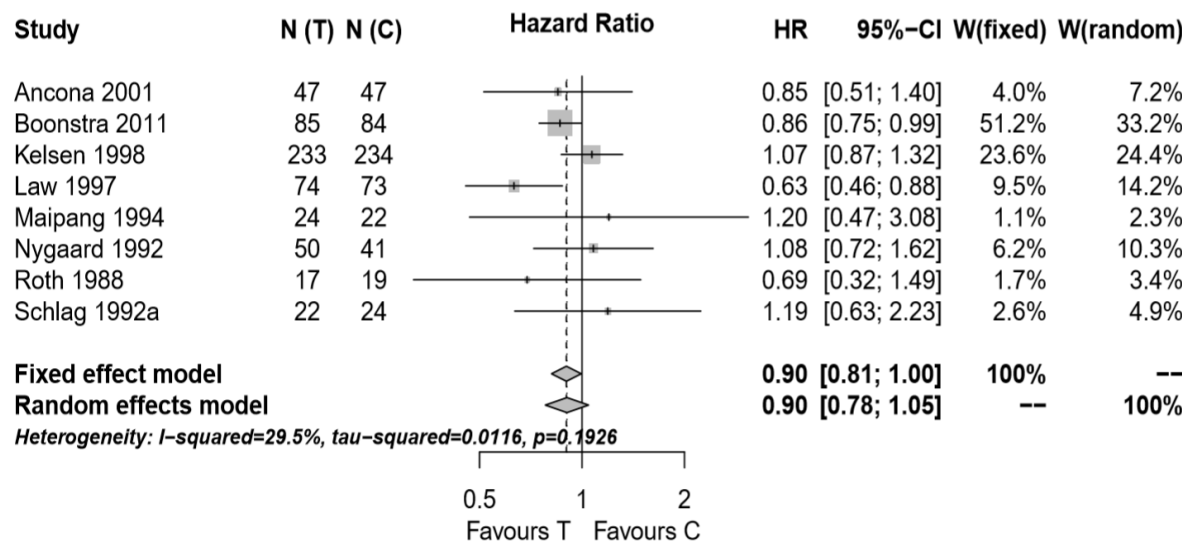


Abbildung 5: Metaanalyse II mit acht Studien nach Ausschluss der Studien Ychou et al. und MRC Allum et al. und Einschluss der Studie Boonstra et al.

Links stehen die acht Studien mit der jeweiligen Patientenanzahl in der Interventions- und Kontrollgruppe. Mittig wird die Hazard Ratio als graues Quadrat und das 95 %-Konfidenzintervall als waagerechte Verlängerung graphisch in einem Forest-Plot dargestellt. Die grauen Rauten symbolisieren die Gesamteffektgröße sowohl unter Annahme eines Festeffektmodells als auch unter Annahme eines Zufallseffektmodells. Rechts werden die Gewichtung der Einzelstudien prozentual sowohl unter Annahme eines Festeffektmodells als auch unter Annahme eines Zufallseffektmodells und die Hazard Ratio mit dem 95 %-Konfidenzintervall numerisch angegeben. Die Heterogenität wird entsprechend ihrer Kennwerte I^2 , τ^2 und p angegeben.

Abkürzungen: C: Kontrollgruppe, HR: Hazard Ratio, $N(T)$: Anzahl der Patienten in der Versuchsgruppe, $N(C)$: Anzahl der Patienten in der Kontrollgruppe, p : P-Wert, T: Versuchsgruppe, $W(fixed)$: Gewicht, das der Studie unter Verwendung eines Festeffektmodells zugewiesen wurde, $W(random)$: Gewicht, das der Studie unter Verwendung eines Zufallseffektmodells zugewiesen wurde

Quelle: Abbildung 4 stammt aus Manzini G, Klotz U, Henne-Bruns D, Kremer M: Validity of studies suggesting preoperative chemotherapy for resectable thoracic esophageal cancer: A critical appraisal of randomized trials. *World Journal of Gastrointestinal Oncology* 12 (2020), S. 120 [48], DOI: 10.4251/wjgo.v12.i1.113 und wurde unter CC BY-NC 4.0 (<http://creativecommons.org/licenses/by-nc/4.0/>) publiziert.

4 Diskussion

In der vorliegenden Arbeit wurde die Validität des Cochrane Reviews und der Metaanalyse mit dem primären Endpunkt *Gesamtüberleben* von *Kidane et al.* [34] anhand einer Validitätsbewertung ihrer zehn zugrundeliegenden Einzelstudien überprüft. Die drei signifikant für eine präoperative Chemotherapie sprechenden Studien mit höchster Gewichtung zeigen aufgrund methodischer Mängel gemäß CONSORT Statement 2010 eine nicht ausreichende Validität und sind somit nur bedingt klinisch aussagekräftig. Nach Ausschluss dieser Studien wird das Ergebnis eines verbesserten Gesamtüberlebens in zwei neu berechneten Metaanalysen nicht mehr unterstützt.

Die Metaanalyse I erfolgte ohne die drei Studien mit höchster Gewichtung (*Boonstra et al.* [10], *MRC Allum et al.* [1] und *Ychou et al.* [84]). Es zeigt sich kein signifikant besseres Gesamtüberleben für Patienten mit präoperativer Chemotherapie bei resektablem thorakalen Ösophaguskarzinom. Dieses Resultat stimmt mit dem ersten Cochrane Review von *Malthaner et al.* aus dem Jahr 2001 und der Metaanalyse von *Urschel et al.* aus dem Jahr 2002 überein [47; 80].

Die Studie von *Boonstra et al.* [10] zeigt gemäß CONSORT Statement 2010 die höchste Validität aller zehn eingeschlossenen Studien. In der Metaanalyse II, in der lediglich die Studien von *MRC Allum et al.* [1] und *Ychou et al.* [84] ausgeschlossen wurden, ergibt sich trotz Einschluss der Studie mit der höchsten Validität, *Boonstra et al.*, kein Überlebensvorteil für Patienten mit präoperativer Chemotherapie.

Mit dem Ziel einer umfassenderen Validitätsanalyse von randomisiert-kontrollierten Studien wurden in dieser Arbeit zusätzlich alle nicht-signifikanten Studien und die signifikante, jedoch niedrig-gewichtete Studie von *Law et al.* [37] überprüft, die in die Metaanalyse der Cochrane Collaboration eingeschlossen wurden. Alle für das positive Ergebnis einer präoperativen Chemotherapie signifikanten Studien weisen keine ausreichende oder zumindest eine fragwürdige Validität auf. Es konnten in mehr als einem Drittel aller CONSORT Kriterien Mängel detektiert werden, obwohl es sich um Studien aus den Jahren 2009 und 2011, die die neuesten zu dieser Thematik sind, handelt. Es ist somit nicht verwunderlich, dass sich in der Untersuchung der nicht-signifikanten Studien, die größtenteils aus den späten 80er und frühen 90er Jahre stammen, noch viel erheblichere Validitätsdefizite zeigen.

Somit sind alle in die Cochrane Metaanalyse eingeschlossenen Studien nicht in der Lage, die aktuellen Standards randomisiert-klinischer Studien zu erfüllen. Im Folgenden sollen die häufigsten Mängel und Kritikpunkte der Studien ausführlicher diskutiert werden.

4.1 Randomisation, Stratifikation und Allocation Concealment

Die Randomisation beschreibt die Methodik, mit der die Probanden einer Studie nach dem Zufallsprinzip den zu vergleichenden Behandlungsgruppen und Kontrollgruppen zugeteilt werden. Sie wird nur in einer Studie, *Ychou et al.* [84], ausreichend beschrieben.

Die elementarste Form zur Generierung einer Randomisierungsliste ist die uneingeschränkte Randomisierung, die sich analog dem wiederholten Werfen einer Münze verhält. Diese Methode vermag es nicht, eine gleichmäßige Verteilung der Probanden in die entsprechenden Behandlungsgruppen oder eine Berücksichtigung bekannter Einflussfaktoren zu gewährleisten.

Um nun unerwünschte Fallzahlunterschiede in den Interventionsgruppen zu vermeiden, kann als Form einer eingeschränkten Randomisation eine sogenannte Blockrandomisation durchgeführt werden. Um weiterhin ein zufälliges Ungleichgewicht bezüglich der Charakteristika in den jeweiligen Forschungsgruppen zu minimieren, kann eine Stratifizierung nach für die Fragestellung relevanter Faktoren (sogenannter *Strata*) wie beispielsweise Alter oder Geschlecht vorgenommen werden. Dies wird auch als geschichtete oder stratifizierte Randomisation bezeichnet, da die Gruppenzuteilung in den einzelnen Strata balanciert wird [68; 64].

Als Alternative zur stratifizierten Randomisation existiert die sogenannte *Minimisation*, die in den Studien von *Ychou et al.* [84] und *MRC Allum et al.* [1] angewandt wurde. Sie wurde als erstes von *Taves* (1974) und *Pocock und Simon* (1975) beschrieben [76; 57] und ist ein Verfahren der adaptiven Randomisierung, die vor allem in onkologischen Studien zunehmend an Popularität gewinnt [83].

Durch die Minimisation werden Gruppen erzeugt, die sowohl hinsichtlich mehreren zuvor festgelegten Strata als auch der Probandenanzahl ausgeglichen sind. Wie der Name bereits impliziert, wird das Ungleichgewicht verschiedener Strata in den Behandlungsgruppen minimiert und sie werden somit zueinander ähnlicher [69].

In der Durchführung wird lediglich der erste Proband einer Gruppe nach dem Zufallsprinzip zugeordnet. Die Allokation aller nachfolgend neu zuzuordnenden Probanden basiert somit auf Informationen der Probanden, die bereits einer Gruppe zugeteilt wurden und richtet sich nach dem aktuellen Gleichgewicht der Strata in Behandlungs- und Kontrollgruppen [44]. Ganz im Sinne einer adaptiven, dynamischen Randomisierung handelt es sich bei der Minimisation um eine Methode, bei der die Gruppenzuordnungswahrscheinlichkeit neuer Probanden von der Kalkulation des potenziellen kovariaten Ungleichgewichts abhängig ist [83].

Jedoch kann die Gruppenzuteilung nicht mehr als rein zufällig angesehen werden, da es sich um ein systematisches Zuteilungsverfahren handelt. Solche Defizite bei der Randomisation führen zu einem Selektionsbias und schließen eine adäquate Geheimhaltung der Randomisierungsliste aus. Gemäß *Berger* (2010) ist eine solche Methodik als äußerst kritisch anzusehen [9]. Dieser lehnt den Vorschlag von *Taves* (2010), dass die Minimisation als Methode der Wahl bei der Zuordnung von Probanden in allen klinischen Studien verwendet werden soll, strikt ab [9; 77].

Darüber hinaus wurde in keiner der zehn Studien eine korrekte Geheimhaltung der Randomisierungsliste durchgeführt. Ein wesentliches Element der Randomisation ist nicht nur die korrekte Erstellung einer Randomisierungsliste, sondern ebenso deren Geheimhaltung. Ohne einwandfreie Anwendung von Maßnahmen zur Geheimhaltung kann es zu Verzerrungen der Studienergebnisse kommen [67].

Studienleiter sollten den genauen Mechanismus der Geheimhaltung nicht nur korrekt umsetzen, sondern diesen im Sinne der CONSORT Checkliste 2010 auch genau beschreiben, um dem Leser die Bewertung ihrer Studie zu vereinfachen.

4.2 Verblindung und Placebo-Kontrolle

Die fehlende Implementierung einer Placebo-Kontrolle und Verblindung beeinträchtigen die Validität aller zehn Studien und zwangsläufig die des darauf aufbauenden Cochrane Reviews.

Ohne Placebo-Kontrolle ist es unmöglich, zwischen spezifischen pharmakologischen oder interventionellen Effekten und Placebo-Effekten zu unterscheiden.

Der Placebo-Effekt ist definiert als Reaktion eines Patienten auf eine Substanz oder ein Verfahren, von dem bekannt ist, dass es keine spezifische therapeutische Wirkung auf die zu behandelnde Erkrankung hat [8].

Mehrere Studien haben gezeigt, dass zum Beispiel die Wahrnehmung von Arzneimitteleigenschaften wie Farbe, Form und Präparationsform, sowie Labortests, die Diagnose und die Arzt-Patienten-Beziehung eine wichtige Rolle für deren Wirkeffekt und somit den Ausgang einer Krankheit spielen [13; 72; 78; 73; 26; 6]. Allein die Information des Patienten der Interventionsgruppe oder Kontrollgruppe anzugehören ist ausreichend, um einen Placebo-Effekt auszulösen [81].

Darüber hinaus könnten auch die Präferenzen, seitens der Patienten oder behandelnden Ärzte, eine Behandlung zu wollen oder nicht, Studienergebnisse beeinflusst haben. So könnten sich beispielsweise Patienten, die der Kontrollgruppe zugeteilt wurden, benachteiligt fühlen, da sie erwartet hatten, therapiert zu werden.

4.3 Fallzahlberechnung

Eine Fallzahlberechnung oder rationale Begründung für den geplanten Stichprobenumfang wurde lediglich in den Studien von *Boonstra et al.* [10], *Kelsen et al.* [33], *Law et al.* [37], *Schlag et al.* [63] und *Ychou et al.* [84] angegeben und fehlt in den übrigen fünf Studien völlig.

Prospektive klinische Studien können nur eine begrenzte Patientenanzahl untersuchen und dienen letztendlich dazu aufgrund der gewonnenen Erkenntnisse aus dem kleineren untersuchten Patientenkollektiv Schlussfolgerungen für eine größere Patientenpopulation zu treffen. Dennoch sollte die begrenzte Patientenanzahl groß genug sein, um mit hoher Wahrscheinlichkeit (*Power*) einen klinisch bedeutsamen, signifikanten Unterschied zwischen der Behandlungs- und Kontrollgruppe zu finden [51].

Demnach ist die Fallzahlplanung ein entscheidender Aspekt im Design einer validen klinischen Studie [60] und ein essentielles Kriterium der CONSORT Checkliste 2010 [51].

4.4 Intention-to-treat-Analyse

Eine Intention-to-treat-Analyse wurde nur in den Studien von *MRC Allum et al.* [1] und *Ychou et al.* [84] durchgeführt.

Dies bedeutet konkret, dass die Messwerte von Probanden, auch wenn sie die ihnen zugedachte Therapie nicht erhalten haben, trotzdem so analysiert werden, als ob sie die zugedachte Therapie erhalten hätten. Demzufolge stellt sie ein wichtiges Verfahren bei der Auswertung einer randomisierten klinischen Studie dar, das die Vergleichbarkeit der Therapiegruppen hinsichtlich bekannter und unbekannter Einflussfaktoren gewährleistet und die Verhältnisse der klinischen Praxis realistischer widerspiegelt [35].

4.5 Ort der Intervention

Der Ort der Interventionen wurde nur in den Studien von *Boonstra et al.* [10] und *Maipang et al.* [46] angegeben. Dieser ist entscheidend, um die Anwendbarkeit und Generalisierbarkeit einer Studie beurteilen zu können.

Gesundheitseinrichtungen unterscheiden sich stark in der Erfahrung des dort tätigen Personals, den verfügbaren Ressourcen oder der internen Organisation. Auch Aspekte des sozioökonomischen und kulturellen Umfelds können sich auf die externe Validität einer Studie auswirken.

Aus diesem Grund fordert das CONSORT Statement 2010 eine genaue Angabe über den Ort der Durchführung und insbesondere, ob eine Studie in einem oder mehreren Zentren (*multizentrische Studie*) durchgeführt wurde. Aus diesen Informationen ist es dem Leser möglich, beurteilen zu können, ob die Ergebnisse einer Studie generalisiert werden und in ihrer eigenen Umgebung relevant sein könnten [51].

4.6 Interessenskonflikte

In der Studie von *MRC Allum et al.* [1] wurden in der Validitätsanalyse finanzielle und nicht-finanzielle Interessenskonflikte (sogenannte *conflicts of interest*) der Studienautoren mit den Sponsoren festgestellt.

Dennis F. Thompson definierte 1993 den Begriff *Interessenskonflikt* im *New England Journal of Medicine* folgendermaßen:

„A conflict of interest is a set of conditions in which professional judgment concerning a primary interest (such as a patient’s welfare or the validity of research) tends to be unduly influenced by a secondary interest (such as financial gain).“

Thompson et al. (1993) S. 573 * [79]

Obwohl in der *Berufsordnung der in Deutschland tätigen Ärztinnen und Ärzte* die Wahrung der ärztlichen Unabhängigkeit ein essenzieller Bestandteil ist [14], sind, wie *Shnier et al.* zeigen, Verbindungen zwischen Autoren und pharmazeutischen Unternehmen häufig und eine potenzielle Quelle für Verzerrungen [70].

Diese Interessenskonflikte machen eine sowohl bewusste als auch unbewusste Manipulation von Methoden und Ergebnissen möglich. Von der pharmazeutischen Industrie geförderte Forschung tendiert eher zu Ergebnissen, die das Produkt des Sponsors begünstigen [41; 71]. Da randomisierten Studien und die daraus hervorgehenden medizinischen Leitlinien maßgeblicher Wert und Vertrauen zugeschrieben werden, sollte der potenzielle Einfluss solcher Interessenskonflikte auf ein Minimum reduziert werden.

Die *Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e. V.* (AWMF) empfiehlt alle direkten und indirekten Interessen der letzten drei Jahre an prominenter Stelle in der Publikation aufzuführen und den Umgang mit etwaig resultierenden Interessenskonflikten zu beschreiben [3]. Obwohl eine vermehrte Transparenz nicht zwangsläufig zur Verminderung von Manipulation führt, stellt sie für die Review-Autoren und interessierte Leser die einzige Möglichkeit dar, sich selbst ein Bild über die Interessenskonflikte der Autoren und somit einem wichtigen

* Thompson D F: Understanding financial conflicts of interest. The New England Journal of Medicine 330: Seite 573 (1993)

Teil der Validität einer Studie zu bilden, so sie denn von den Autoren überhaupt als solche deklariert wurden.

4.7 Limitationen

Eine mögliche Limitation dieser Arbeit ist die Verwendung des CONSORT Statements 2010 zur Validitätsbewertung. Da es sich um eine Leitlinie für Forschungsberichte (sogenannte *Reporting Guideline*) handelt, ist sie ursprünglich nicht ein Werkzeug für die Qualitätsbeurteilung von klinischen Studien.

Um Studien auf mögliche Fehlerquellen zu untersuchen und das Risiko systematischer Fehler (sogenannter *risk of bias*) zu bewerten, stehen eine Vielzahl von Skalen, Checklisten und Fragebögen zur Verfügung [21]. Auf empirischen Forschungsergebnissen evaluiert wurde beispielsweise das *Risk of Bias Tool* der Cochrane Collaboration. Dies ist ein Komponentensystem, in dem das Verzerrungspotenzial für verschiedene Bereiche wie Randomisierung, Geheimhaltung der Gruppenzuteilung, Verblindung und Daten separat berichtet und bewertet wird [28].

Das CONSORT Statement 2010 enthält keine Empfehlung zu Design oder Analyse von Studien. Es beinhaltet evidenzbasierte Mindeststandards zur Durchführung und Dokumentation einer randomisiert-kontrollierten Studie, um die Qualität und Plausibilität ihrer Berichterstattung zu verbessern und eine transparente Berichterstattung zu erleichtern. Weltweit wird das CONSORT Statement von zahlreichen Fachzeitschriften wie *The Lancet*[®], *BMC*[®], *BMJ*[®], *JAMA*[®] oder *The New England Journal of Medicine*[®] befürwortet [17]. Ein im Jahr 2006 veröffentlichtes systematisches Review konstatiert seit ihrer Einführung eine verbesserte Berichterstattung von randomisiert-kontrollierten Studien [56].

Es ist auch ein positiver, indirekter Einfluss auf die methodische Qualität der durchzuführenden Studie anzunehmen. Gewissenhafte und vollständige Berichterstattung sollten es praktisch unmöglich machen, dass methodisch ungenügende Studien den Publikationsprozess durchlaufen, ohne dass dabei Mängel bei der Studiendurchführung aufgedeckt werden [55].

Zudem konstatiert *Meerpohl et al.*, dass Leitlinien für Forschungsberichte auch für Leser medizinischer Fachartikel hilfreich sein können, um sich über die methodische Qualität einer Studie kundig zu machen [50]. Dem zu folgend kann die Verwendung

der CONSORT Checkliste des CONSORT Statements 2010 als Tool für die Beurteilung der Validität von Studien als geeignet angenommen werden.

4.8 Schlussfolgerungen

Die Untersuchung dieser Arbeit deckt mithilfe des CONSORT Statements 2010 zu viele mögliche systematische Verzerrungen der für das Ergebnis einer Befürwortung einer präoperativen Chemotherapie relevanten Studien von *Boonstra et al.* [10], *MRC Allum et al.* [1] und *Ychou et al.* [84] auf, sodass diese als nicht ausreichend valide genug und nur bedingt klinisch aussagekräftig einzustufen sind. Sie gefährden die Validität der Metaanalyse des Cochrane Reviews von *Kidane et al.* [34] in entscheidender Weise und hätten nicht miteingeschlossen werden dürfen. Bei Ausschluss dieser drei Studien zeigen zwei neu durchgeführte Metaanalysen kein signifikant verbessertes Gesamtüberleben mehr.

Die übrigen sechs nicht-signifikanten Studien (*Ancona et al.* [2], *Kelsen et al.* [33], *Maipang et al.* [46], *Nygaard et al.* [53], *Roth et al.* [61] und *Schlag et al.* [63]) und die signifikante, jedoch von den Review-Autoren niedrig gewichtete Studie von *Law et al.* [37] wurden ebenfalls auf Validität geprüft. Alle wurden vor dem Jahr 2001 durchgeführt. Drei Studien stammen aus den frühen 90er Jahre (*Maipang et al.* (1994) [46], *Nygaard et al.* (1992) [53] und *Schlag et al.* (1992) [63]), eine Studie reicht sogar bis in die 80er Jahre zurück (*Roth et al.* (1988) [61]). Zu dieser Zeit existierten noch keine Empfehlungen für die Erstellung von Studienberichten für randomisiert-kontrollierte Studien. Das CONSORT Statement 2010 wurde erstmals 1996 als erste Leitlinie für die transparente Berichterstattung von randomisierten, klinischen Studien veröffentlicht und jeweils in den Jahren 2001 und 2010 aktualisiert [7; 52; 66]. Diese Tatsache spiegelt sich auch in den Untersuchungsergebnissen wider: Die älteren Studien weisen im Vergleich zu den neueren Studien eine noch viel geringere Validität gemäß CONSORT Statement 2010 auf und würden nach heutigen Qualitätsstandards die Mindeststandards für eine Publikation in einem *peer-reviewed* Journal nicht mehr erfüllen.

Eine solch mangelhafte Durchführung und Beschreibung von klinischen Studien verdeutlicht nur einmal mehr die absolute Notwendigkeit des CONSORT Statements 2010 und vergleichbarer Qualitätsinstrumente.

Diese Arbeit soll folglich nicht nur das Ergebnis des Cochrane Reviews von *Kidane et al.* [34] kritisch hinterfragen, sondern auch das übergeordnete Ziel haben, Studienautoren für methodisch besser durchgeführte Studien und vollständige Dokumentation zu sensibilisieren.

Bei insgesamt niedriger Inzidenz des Ösophaguskarzinoms wurden Adenokarzinome und Plattenepithelkarzinome ohne weitere Trennung der histopathologischen Subtypen voneinander in die Studien eingeschlossen. Nicht nur unterschiedliche prognostische Implikationen beider histopathologischer Karzinomtypen, sondern auch die häufig durchgeführte Kombinationstherapie aus Chemotherapie und Radiotherapie zu unterschiedlichen Therapiezeitpunkten können zu einer Verzerrung der Studienergebnisse beigetragen haben. So erhielten in der Studie von *Ychou et al.* [84] 50 % der Patienten der Interventionsgruppe bei guter Toleranz und fehlendem Hinweis auf Tumorprogress bis zu 4 Zyklen postoperativer Chemotherapie mit 5-Fluoruracil und Cisplatin. In der Studie von *MRC Allum et al.* [1] erhielten 9 % der Patienten sowohl in der Interventions- als auch Kontrollgruppe eine präoperative Bestrahlung mit 25 Gy fraktioniert über eine Woche, 32,5 Gy fraktioniert über zwei Wochen oder einer äquivalenten Dosis.

Folglich ist es nicht mehr möglich nachzuvollziehen, auf welches Therapiekonzept und bei welchem histopathologischen Subtypen der gemessene Behandlungseffekt tatsächlich zurückzuführen ist. Die Autoren des Cochrane Reviews von *Kidane et al.* [34] legen in ihren Methoden fest, dass Studien mit anderen Therapiemodalitäten als einer präoperativen Chemotherapie ausgeschlossen wurden:

„We excluded studies if they were not truly randomized, earlier versions of updated trials reporting on the same participants, if other treatment modalities (eg, radiotherapy, hyperthermia) were used, or if no surgery alone control arms were included.“
Kidane et al. (2015) S. 6 * [34].

* Kidane et al.: Preoperative chemotherapy for resectable thoracic esophageal cancer (Review). The Cochrane Database of Systematic Reviews 5: Seite 6 (2015)

Demnach sind die Studien von *Ychou et al.* [84] und *MRC Allum et al.* [1], die dieses Kriterium nicht erfüllt haben, für das Cochrane Review nicht geeignet gewesen.

Um die Rolle einer präoperativen Chemotherapie zu untersuchen, müssten demnach zwei Gruppen gebildet werden: *Radiotherapie + Operation* versus *Radiochemotherapie + Operation*, wie es von *Herskovic et al.* durchgeführt wurde [27]. Es wurde jedoch auch in dieser Studie keine Trennung zwischen Adeno- und Plattenepithelkarzinomen vorgenommen.

Fazit

Verschiedene nationale Leitlinien empfehlen heutzutage ein multimodales Therapieregime bei der Therapie des Ösophaguskarzinom [45]. In Deutschland wurde die S3-Leitlinie zur Diagnostik und Therapie des Ösophaguskarzinoms zuletzt im Jahr 2018 aktualisiert [20]. Dabei wurden auch die Studien von *Boonstra et al.* [10], *MRC Allum et al.* [1] und *Ychou et al.* [84] berücksichtigt. Deren Studienergebnisse werden von den Leitlinienautoren jedoch nur bedingt in die Therapieempfehlungen der Leitlinie aufgenommen:

„Bei Plattenepithelkarzinomen lässt sich ein konsistenter Überlebenszugewinn trotz der positiven Studie von Boonstra in Metaanalysen nicht belegen.“

S3-Leitlinie Diagnostik und Therapie der Plattenepithelkarzinome und Adenokarzinome des Ösophagus (2018) S. 101* [20].

Die Metaanalyse des Cochrane Reviews von *Kidane et al.* [34] findet in dieser Leitlinie hingegen keine Erwähnung [20].

Für die Beantwortung der Frage nach der für den Patienten besten multimodalen Therapie bei resektablem thorakalen Ösophaguskarzinom werden künftig qualitativ hochwertige multizentrisch randomisiert-kontrollierte Studien durchzuführen sein, die zusammenfassend in einer Metaanalyse bewertet werden sollten.

Denn nur hohe Anforderungen an die methodische Studienqualität und die durch vollständige Dokumentation gewährleistete Validität, ist vor dem Hintergrund einer evidenzbasierten Medizin das Fundament einer jeden klinischen Behandlungsempfehlung.

* Deutsche Krebsgesellschaft, Deutsche Krebshilfe, AWMF (Leitlinienprogramm Onkologie): S3-Leitlinie Diagnostik und Therapie der Plattenepithelkarzinome und Adenokarzinome des Ösophagus. AWMF online Langversion 2.0: Seite 101 (2018)

5 Zusammenfassung

Seit Entwicklung der chirurgischen Resektion zu Beginn des 20. Jahrhunderts stellt diese die einzige potenziell kurative Behandlungsform des thorakalen Ösophaguskarzinoms dar und ist somit der Goldstandard der Therapie. Aufgrund hoher Rezidiv- und geringer Überlebensraten wurde in den letzten drei Jahrzehnten durch zusätzliche Therapiemodalitäten ein multidisziplinäres Therapieregime zu etablieren versucht. Ein Cochrane Review aus dem Jahr 2015 zeigt in einer Metaanalyse eine signifikante Verlängerung des Gesamtüberlebens für Patienten mit präoperativer Chemotherapie bei resektablem thorakalen Ösophaguskarzinom. Ziel dieser Arbeit ist die Untersuchung der Validität der Metaanalyse dieses Cochrane Reviews anhand einer Validitätsbewertung der zugrundeliegenden Einzelstudien. Insbesondere soll hierbei die Validität der signifikant für eine präoperative Chemotherapie sprechenden Studien geprüft und anschließend die Aussagekraft der Metaanalyse erneut bewertet.

Das Hauptzielkriterium *Gesamtüberleben* wurde von den Autoren des Cochrane Reviews mit einer zehn Einzelstudien umfassenden Metaanalyse bewertet. Drei Studien waren sowohl signifikant als auch hochgewichtet und damit ausschlaggebend für die Empfehlung einer präoperativen Chemotherapie. Die Studienqualität aller Einzelstudien wurde mithilfe des evidenzbasierten CONSORT Statements 2010 untersucht. Es wurden zwei neue Metaanalysen durchgeführt: Metaanalyse I wurde ohne die drei maßgeblich für eine präoperative Chemotherapie sprechenden Studien berechnet. Metaanalyse II schloss eine Studie ein, die gemäß CONSORT Statement 2010 die höchste Validität der drei vorgenannten Studien besitzt. In beide neuen Metaanalysen wurden die übrigen sieben Studien eingeschlossen.

Die signifikant für eine präoperative Chemotherapie sprechenden Studien mit hoher Gewichtung erfüllen über ein Drittel der CONSORT Kriterien nicht. Die übrigen Studien zeigen für mehr als 40 % der CONSORT Kriterien keine ausreichende Validität. Beide neu berechneten Metaanalysen zeigen statistisch keinen signifikanten Überlebensvorteil mehr.

Für die signifikant für eine präoperative Chemotherapie sprechenden Studien lassen sich durch fehlende oder systematische Randomisierung ein Selektionsbias nicht ausschließen und ein adäquates Allocation Concealment nicht umsetzen.

Fehlende Verblindung und Placebo-Kontrollen machen eine Unterscheidung zwischen spezifischen pharmakologischen und Placebo-Effekten unmöglich. Potenziell systematische Verzerrungen entstehen darüber hinaus durch unzureichende Darstellung der Berechnungsgrundlage, der Stichprobengrößen sowie Confounder durch Kombinationstherapien aus präoperativer Chemotherapie und Radiatio sowie zusätzlichen postoperativen Chemotherapien. Somit sind diese Studien durch methodische Mängel gemäß CONSORT Statement nicht ausreichend valide, um in einer Metaanalyse berücksichtigt zu werden.

Unter konsequenter Herausnahme dieser Studien zeigt Metaanalyse I kein signifikant verbessertes Gesamtüberleben mehr. Auch unter Einschluss der validesten Studie erbringt Metaanalyse II kein signifikant verbessertes Gesamtüberleben mehr. In Zusammenschau ist die Empfehlung einer präoperativen Chemotherapie bei resektablem thorakalen Ösophaguskarzinom des Cochrane Reviews als nicht ausreichend evidenzbasiert zu betrachten.

Auch die Analyse der übrigen Studien zeigt gravierende Validitätsdefizite. Somit sind alle in die Cochrane Metaanalyse eingeschlossenen Studien nicht in der Lage, die aktuellen Standards randomisiert-klinischer Studien zu erfüllen.

Die drei signifikant für eine präoperative Chemotherapie sprechenden Studien werden in der deutschen S3-Leitlinie zur Diagnostik und Therapie des Ösophaguskarzinoms von 2018 berücksichtigt, jedoch kritisch hinterfragt und deren Studienergebnisse nur bedingt in die Therapieempfehlung der Leitlinie aufgenommen. Die Metaanalyse des Cochrane Reviews findet in dieser Leitlinie keine Erwähnung. Für die Entscheidung der bestmöglichen multimodalen Therapie bei resektablem thorakalen Ösophaguskarzinom werden künftig qualitativ hochwertige multizentrisch randomisiert-kontrollierte Studien durchzuführen sein. Besonderer Stellenwert muss auf methodischer Validität und vollständiger Dokumentation liegen, die den aktuellen Qualitätsstandards randomisierter-kontrollierter Studien und evidenzbasierter Medizin entsprechen.

6 Literaturverzeichnis

- [1] Allum W H, Stenning S P, Bancewicz J, Clark P I, Langley R E: Long-term results of a randomized trial of surgery with or without preoperative chemotherapy in esophageal cancer. *Journal of Clinical Oncology* 27: 5062–5067 (2009)
- [2] Ancona E, Ruol A, Santi S, Merigliano S, Sileni V C, Koussis H, Zaninotto G, Bonavina L, Peracchia A: Only pathologic complete response to neoadjuvant chemotherapy improves significantly the long term survival of patients with resectable esophageal squamous cell carcinoma: final report of a randomized, controlled trial of preoperative chemotherapy versus surgery alone. *Cancer* 91: 2165–2174 (2001)
- [3] Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V. (AWMF): Empfehlungen der AWMF zum Umgang mit Interessenkonflikten bei Aktivitäten wissenschaftlicher medizinischer Fachgesellschaften – Fassung vom 10. November 2017 https://www.awmf.org/fileadmin/user_upload/Leitlinien/AWMF-Publikationen/20171110_AWMF_Empfehlungen_zu_Interessenkonflikten__V2.2_f.pdf (12.06.2021)
- [4] Baba M, Natsugoe S, Shimada M, Nakano S, Kusano C, Fukumoto T, Aikou T, Akazawa K: Prospective evaluation of preoperative chemotherapy in resectable squamous cell carcinoma of the thoracic esophagus. *Official Journal of the International Society for Diseases of the Esophagus* 13: 136–141 (2000)
- [5] Barraclough H, Simms L, Govindan R: Biostatistics primer: what a clinician ought to know: hazard ratios. *Journal of Thoracic Oncology* 6: 978–982 (2011)
- [6] Bass M J, Buck C, Turner L, Dickie G, Pratt G, Robinson H C: The physician's actions and the outcome of illness in family practice. *The Journal of Family Practice* 23: 43–47 (1986)
- [7] Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz K F, Simel D, Stroup D F: Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *The Journal of the American Medical Association* 276: 637–639 (1996)

- [8] Benedetti F, Amanzio M: The neurobiology of placebo analgesia: from endogenous opioids to cholecystokinin. *Progress in Neurobiology* 52: 109–125 (1997)
- [9] Berger V W: Minimization, by its nature, precludes allocation concealment, and invites selection bias. *Contemporary Clinical Trials* 31: 406 (2010, Brief an den Herausgeber)
- [10] Boonstra J J, Kok T C, Wijnhoven B P, van Heijl M, van Berge Henegouwen M I, Kate F J ten, Siersema P D, Dinjens W N, van Lanschot J J, Tilanus H W, van der Gaast A: Chemotherapy followed by surgery versus surgery alone in patients with resectable oesophageal squamous cell carcinoma: long-term results of a randomized controlled trial. *BioMed Central Cancer* 11: 1–10 (2011)
- [11] Borenstein M, Hedges L V, Higgins J P T, Rothstein H R: Part 3: Fixed-Effect Versus Random-Effects Models. In: Borenstein M, Hedges L V, Higgins J P T, Rothstein H R (Hrsg) *Introduction to Meta-Analysis*, 1. Edition, John Wiley & Sons, Ltd: 61-102, Chichester (2009)
- [12] Bown M J, Sutton A J: Quality control in systematic reviews and meta-analyses. *European Journal of Vascular and Endovascular Surgery* 40: 669–677 (2010)
- [13] Buckalew L W, Coffield K E: An investigation of drug expectancy as a function of capsule color and size and preparation form. *Journal of Clinical Psychopharmacology* 2: 245–248 (1982)
- [14] Bundesärztekammer: (Muster-)Berufsordnung für die in Deutschland tätigen Ärztinnen und Ärzte (Stand 2018) <https://www.bundesaerztekammer.de/recht/berufsrecht/muster-berufsordnung-aerzte/muster-berufsordnung/> (06.06.2021)
- [15] Chenot J-F, Reber K C: Metaanalysen lesen und interpretieren: eine praktische Anleitung. *Zeitschrift für Allgemeinmedizin* 11: 469–473 (2015)
- [16] Cochrane Collaboration: About us <https://www.cochrane.org/about-us> (07.06.2021)
- [17] CONSORT: Endorsers, Journals and Organizations <http://www.consort-statement.org/about-consort/endorsers1> (11.06.2021)

- [18] Deeks, J J, Higgins, J P, Altman, D G: Chapter 10: Analysing data and undertaking meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (Hrsg) Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021) <https://training.cochrane.org/handbook/current/chapter-10#section-10-10> (07.06.2021)
- [19] DerSimonian R, Laird N: Meta-analysis in clinical trials. *Controlled Clinical Trials* 7: 177–188 (1986)
- [20] Deutsche Krebsgesellschaft, Deutsche Krebshilfe, AWMF (Leitlinienprogramm Onkologie): S3-Leitlinie Diagnostik und Therapie der Plattenepithelkarzinome und Adeno-karzinome des Ösophagus, Langversion 2.0, 2018 https://www.leitlinienprogramm-onkologie.de/fileadmin/user_upload/Downloads/Leitlinien/O-esophaguskarzinom/Version_2/LL_Oesophagus_Langversion_2.0.pdf (27.02.2020)
- [21] Dreier M, Borutta B, Stahmeyer J, Krauth C, Walter U: Comparison of tools for assessing the methodological quality of primary and secondary studies in health technology assessment reports in Germany. *GMS Health Technology Assessment* 6: 1–11 (2010)
- [22] Du Prel J-B, Hommel G, Röhrig B, Blettner M: Confidence interval or p-value?: part 4 of a series on evaluation of scientific publications. *Deutsches Ärzteblatt International* 106: 335–339 (2009)
- [23] Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin D M, Forman D, Bray F: Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer* 136: 359-386 (2015)
- [24] Garg P K, Sharma J, Jakhetiya A, Goel A, Gaur M K: Preoperative therapy in locally advanced esophageal cancer. *World Journal of Gastroenterology* 22: 8750–8759 (2016)
- [25] Glass G V: Primary, Secondary, and Meta-Analysis of Research. *Educational Researcher* 5: 3–8 (1976)
- [26] Greenfield S, Kaplan S, Ware J E: Expanding patient involvement in care. Effects on patient outcomes. *Annals of Internal Medicine* 102: 520–528 (1985)

- [27] Herskovic A, Martz K, al-Sarraf M, Leichman L, Brindle J, Vaitkevicius V, Cooper J, Byhardt R, Davis L, Emami B: Combined chemotherapy and radiotherapy compared with radiotherapy alone in patients with cancer of the esophagus. *The New England Journal of Medicine* 326: 1593–1598 (1992)
- [28] Higgins J P T, Altman D G, Gøtzsche P C, Jüni P, Moher D, Oxman A D, Savovic J, Schulz K F, Weeks L, Sterne J A C: The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *British Medical Journal* 343: 1-9 (2011)
- [29] Higgins J P T, Thompson S G, Deeks J J, Altman D G: Measuring inconsistency in meta-analyses. *British Medical Journal* 327: 557–560 (2003)
- [30] Jüni P, Altman D G, Egger M: Systematic reviews in health care: Assessing the quality of controlled clinical trials. *British Medical Journal* 323: 42–46 (2001)
- [31] Kabisch M, Ruckes C, Seibert-Grafe M, Blettner M: Randomized controlled trials: part 17 of a series on evaluation of scientific publications. *Deutsches Ärzteblatt International* 108: 663–668 (2011)
- [32] Kreditsu K K, Jiwnani S, Karimundackal G, Pramesh C S: Multimodality management of esophageal cancer. *Indian Journal of Surgical Oncology* 4: 96–104 (2013)
- [33] Kelsen D P, Ginsberg R, Pajak T F, Sheahan D G, Gunderson L, Mortimer J, Estes N, Haller D G, Ajani J, Kocha W, Minsky B D, Roth J A: Chemotherapy followed by surgery compared with surgery alone for localized esophageal cancer. *The New England Journal of Medicine* 339: 1979–1984 (1998)
- [34] Kidane B, Coughlin S, Vogt K, Malthaner R: Preoperative chemotherapy for resectable thoracic esophageal cancer. *The Cochrane Database of Systematic Reviews* 5: 1-50 (2015)
- [35] Kleist P: Das Intention-to-Treat-Prinzip. *Swiss Medical Forum* 9: 450–454 (2009)
- [36] Knippschild S, Baulig C, Krummenauer F: Heterogenität in Meta-Analysen – kein Vergleich von Äpfeln und Birnen erlaubt ... *Zeitschrift für Zahnärztliche Implantologie* 31: 224–229 (2015)

- [37] Law S, Fok M, Chow S, Chu K-M, Wong J: Preoperative chemotherapy versus surgical therapy alone for squamous cell carcinoma of the esophagus: A prospective randomized trial. *The Journal of Thoracic and Cardiovascular Surgery* 114: 210–217 (1997)
- [38] Law S Y, Fok M, Cheng S, Wong J: A comparison of outcome after resection for squamous cell carcinomas and adenocarcinomas of the esophagus and cardia. *Surgery, Gynecology & Obstetrics* 175: 107–112 (1992)
- [39] Lerut T: Esophageal surgery at the end of the millennium. *The Journal of Thoracic and Cardiovascular Surgery* 116: 1–20 (1998)
- [40] Lerut T, Leyn P de, Coosemans W, van Raemdonck D, Scheys I, LeSaffre E: Surgical strategies in esophageal carcinoma with emphasis on radical lymphadenectomy. *Annals of Surgery* 216: 583–590 (1992)
- [41] Lexchin J, Bero L A, Djulbegovic B, Clark O: Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *British Medical Journal* 326: 1167–1170 (2003)
- [42] Li F, Ding N, Zhao Y, Yuan L, Mao Y: The current optimal multimodality treatments for oesophageal squamous-cell carcinoma: A systematic review and meta-analysis. *International Journal of Surgery* 60: 88–100 (2018)
- [43] Lieberman M D, Shriver C D, Bleckner S, Burt M: Carcinoma of the esophagus: Prognostic significance of histologic type. *The Journal of Thoracic and Cardiovascular Surgery* 109: 130–139 (1995)
- [44] Lim C-Y, In J: Randomization in clinical studies. *Korean Journal of Anesthesiology* 72: 221–232 (2019)
- [45] M. Martin-Richard, R. Díaz Beveridge, V. Arrazubi, M. Alsina, M. Galan Guzmán, A. B. Custodio, C. Gómez, F. L. Muñoz, R. Pazo, F. Rivera: SEOM Clinical Guideline for the diagnosis and treatment of esophageal cancer (2016). *Clinical and Translational Oncology* 18: 1179–1186 (2016)
- [46] Maipang T, Vasinanukorn P, Petpichetchian C, Chamroonkul S, Geater A, Chansawwaang S, Kuapanich R, Panjapiyakul C, Watanaarepornchai S, Punnperk S: Induction chemotherapy in the treatment of patients with carcinoma of the esophagus. *Journal of Surgical Oncology* 56: 191–197 (1994)

- [47] Malthaner R A, Collin S, Fenlon D: Preoperative chemotherapy for resectable thoracic esophageal cancer. The Cochrane Database of Systematic Reviews 1: 1-47 (2001)
- [48] Manzini G, Klotz U, Henne-Bruns D, Kremer M: Validity of studies suggesting preoperative chemotherapy for resectable thoracic esophageal cancer: A critical appraisal of randomized trials. World Journal of Gastrointestinal Oncology 12: 113–123 (2020)
- [49] Medical Research Council Oesophageal Cancer Working Group: Surgical resection with or without preoperative chemotherapy in oesophageal cancer: a randomised controlled trial. The Lancet 359: 1727–1733 (2002)
- [50] Meerpohl J J, Blümle A, Antes G, von Elm E: Leitlinien für Forschungsberichte sind auch für Leser medizinischer Fachartikel hilfreich: CONSORT, STARD, STROBE & Co. Deutsche Medizinische Wochenschrift 134: 2078–2083 (2009)
- [51] Moher D, Hopewell S, Schulz K F, Montori V, Gøtzsche P C, Devereaux P J, Elbourne D, Egger M, Altman D G: CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. British Medical Journal 340: 1-28 (2010)
- [52] Moher D, Schulz K, Altman D: Das CONSORT Statement: Überarbeitete Empfehlungen zur Qualitätsverbesserung von Reports randomisierter Studien im Parallel-Design. Deutsche Medizinische Wochenschrift 129: 16-20 (2004)
- [53] Nygaard K, Hagen S, Hansen H S, Hatlevoll R, Hultborn R, Jakobsen A, Mäntylä M, Modig H, Munck-Wikland E, Rosengren B: Pre-operative radiotherapy prolongs survival in operable esophageal carcinoma: a randomized, multicenter study of pre-operative radiotherapy and chemotherapy. The second Scandinavian trial in esophageal cancer. World Journal of Surgery 16: 1104-1109 (1992)
- [54] Orringer M B: Multimodality therapy for esophageal carcinoma--update. Chest 103: 406–409 (1993)
- [55] Pittler M, Blümle A, Meerpohl J, Antes G: CONSORT 2010: Aktualisierte Leitlinie für Berichte randomisierter Studien im Parallelgruppen-Design. Deutsche Medizinische Wochenschrift 136: 20-23 (2011)

- [56] Plint A C, Moher D, Morrison A, Schulz K, Altman D G, Hill C, Gaboury I: Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *The Medical Journal of Australia* 185: 263–267 (2006)
- [57] Pocock S J, Simon R: Sequential Treatment Assignment with Balancing for Prognostic Factors in the Controlled Clinical Trial. *Biometrics* 31: 103–115 (1975)
- [58] Posner M C, Minsky B D, Ilson D H: Chapter 45: Cancer of the Esophagus. In: DeVita V T, Lawrence T S, Rosenberg S A (Hrsg) *DeVita, Hellman, and Rosenberg's Cancer: Principles & Practice of Oncology*, 10. Edition, Wolters Kluwer: 574-612, Philadelphia (2015)
- [59] Robert Koch Institut, Gesellschaft der epidemiologischen Krebsregister in Deutschland: Krebs in Deutschland 2015/2016 https://www.krebsdaten.de/Krebs/DE/Content/Publikationen/Krebs_in_Deutschland/kid_2019/krebs_in_deutschland_2019.pdf?__blob=publicationFile (09.06.2021)
- [60] Röhrig B, Du Prel J-B, Wachtlin D, Kwiecien R, Blettner M: Sample size calculation in clinical trials: part 13 of a series on evaluation of scientific publications. *Deutsches Ärzteblatt International* 107: 552–556 (2010)
- [61] Roth J A, Pass H I, Flanagan M M, Graeber G M, Rosenberg J C, Steinberg S: Randomized clinical trial of preoperative and postoperative adjuvant chemotherapy with cisplatin, vindesine, and bleomycin for carcinoma of the esophagus. *The Journal of Thoracic and Cardiovascular Surgery* 96: 242–248 (1988)
- [62] Sackett D L, Rosenberg W M, Gray J A, Haynes R B, Richardson W S: Evidence based medicine: what it is and what it isn't. *British Medical Journal* 312: 71–72 (1996)
- [63] Schlag P M: Randomized trial of preoperative chemotherapy for squamous cell cancer of the esophagus. The Chirurgische Arbeitsgemeinschaft Fuer Onkologie der Deutschen Gesellschaft Fuer Chirurgie Study Group. *Archives of Surgery* 127: 1446–1450 (1992)

- [64] Schulz K F: Randomized controlled trials. *Clinical Obstetrics and Gynecology* 41: 245–256 (1998)
- [65] Schulz K F, Altman D G, Moher D: CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *PLoS Medicine* 7: 1-7 (2010)
- [66] Schulz K F, Altman D G, Moher D: CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *British Medical Journal* 340: 1-23 (2010)
- [67] Schulz K F, Chalmers I, Hayes R J, Altman D G: Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *The Journal of the American Medical Association* 273: 408–412 (1995)
- [68] Schulz K F, Grimes D A: Reihe Epidemiologi 6: Generierung von Randomisierungslisten in randomisierten Studien: Zufall, nicht Auswahl. *German Journal for Quality in Health Care* 101: 419–426 (2007)
- [69] Scott N W, McPherson G C, Ramsay C R, Campbell M K: The method of minimization for allocation to clinical trials. *Controlled Clinical Trials* 23: 662–674 (2002)
- [70] Shnier A, Lexchin J, Romero M, Brown K: Reporting of financial conflicts of interest in clinical practice guidelines: a case study analysis of guidelines from the Canadian Medical Association Infobase. *BioMed Central Health Services Research* 16: 1–8 (2016)
- [71] Sismondo S: Pharmaceutical company funding and its consequences: a qualitative systematic review. *Contemporary Clinical Trials* 29: 109–113 (2008)
- [72] Sox H C, Margulies I, Sox C H: Psychologically mediated effects of diagnostic tests. *Annals of Internal Medicine* 95: 680–685 (1981)
- [73] Stewart M A: Effective physician-patient communication and health outcomes: a review. *Canadian Medical Association Journal* 152: 1423–1433 (1995)

- [74] Stilidi I, Bokhyan V, Tryakin A, Suleymanov E, Kononets P, Malikhova O, Tjulandin S: Preoperative chemotherapy followed by resection vs. surgery alone for locally advanced esophageal carcinoma: Interim analysis of a randomized study. *Journal of Clinical Oncology*: 4055 (2006, Abstrakt)
- [75] Sung H, Ferlay J, Siegel R L, Laversanne M, Soerjomataram I, Jemal A, Bray F: Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *A cancer Journal for Clinicians* 71: 209–249 (2021)
- [76] Taves D R: Minimization: a new method of assigning patients to treatment and control groups. *Clinical Pharmacology and Therapeutics* 15: 443–453 (1974)
- [77] Taves D R: The use of minimization in clinical trials. *Contemporary Clinical Trials* 31: 180–184 (2010)
- [78] Thomas K B: General practice consultations: is there any point in being positive? *British Medical Journal* 294: 1200–1202 (1987)
- [79] Thompson D F: Understanding financial conflicts of interest. *The New England Journal of Medicine* 329: 573–576 (1993)
- [80] Urschel J D, Vasan H, Blewett C J: A meta-analysis of randomized controlled trials that compared neoadjuvant chemotherapy and surgery to surgery alone for resectable esophageal cancer. *The American Journal of Surgery* 183: 274–279 (2002)
- [81] Waber R L, Shiv B, Carmon Z, Ariely D: Commercial features of placebo and therapeutic efficacy. *The Journal of the American Medical Association* 299: 1016–1017 (2008)
- [82] Wang C, Ding T, Chang L: A randomized clinical study of preoperative chemotherapy for esophageal carcinoma. *Chinese Journal of Oncology* 23: 254–255 (2001)
- [83] Xu Z, Proschan M, Lee S: Validity and power considerations on hypothesis testing under minimization. *Statistics in Medicine* 35: 2315–2327 (2016)

- [84] Ychou M, Boige V, Pignon J-P, Conroy T, Bouché O, Lebreton G, Ducourtieux M, Bedenne L, Fabre J-M, Saint-Aubert B, Genève J, Lasser P, Rougier P: Perioperative chemotherapy compared with surgery alone for resectable gastroesophageal adenocarcinoma: an FNCLCC and FFCD multicenter phase III trial. *Journal of Clinical Oncology* 29: 1715–1721 (2011)
- [85] Zwiener I, Blettner M, Hommel G: Survival Analysis - part 15 of a series on evaluation of scientific publications. *Deutsches Ärzteblatt International* 108: 163-169 (2011)

Danksagungen

Mein großer Dank gebührt zunächst der zwischenzeitlich emeritierten ärztlichen Direktorin für Allgemein- und Viszeralchirurgie des Universitätsklinikums Ulm, Frau Prof. Dr. med. Doris Henne-Bruns, für die vertrauensvolle Überlassung des Themas.

Mein außerordentlicher Dank gilt Herrn PD Dr. med. Michael Kremer, Leitender Arzt der Viszeralchirurgie am Kantonsspital Aarau, für die Übernahme der fachlichen Betreuung als Doktorvater, der zuteilgewordenen Unterstützung und Beantwortung jeweiliger offener Fragen in freundlichen und bereichernden Gesprächen.

Des Weiteren möchte ich mich herzlichst bei Frau PD Dr. med. Giulia Manzini, Oberärztin der Viszeralchirurgie am Kantonsspital Aarau, für die exzellente statistische Beratung bedanken.

Für die Durchsicht und ausgesprochen wertvollen Anregungen möchte ich mich bei Oberstabsarzt Conn Rother ausdrücklich bedanken.

Mein ganz besonderer Dank richtet sich an meine Eltern und meinen Bruder sowie den individuellen Lebensumständen, die mir das Studium der Humanmedizin ermöglichten.

Lebenslauf

Der Lebenslauf wurde aus Gründen des Datenschutzes entfernt.