

**MS-Therapiestudien,
evaluiert nach CONSORT und
Evidence-based Medicine
– Neuer Beurteilungsscore für die
Validität und Review –**

Christian Hassler

Institut für Medizinische Statistik und Epidemiologie
der Technischen Universität München
Klinikum rechts der Isar
(Direktor: Univ.-Prof. Dr. A. Neiß)

MS-Therapiestudien, evaluiert nach CONSORT und Evidence-based Medicine – Neuer Beurteilungsscore für die Validität und Review –

Christian Hassler

Vollständiger Abdruck der von der Fakultät für Medizin
der Technischen Universität München zur Erlangung des akademischen Grades eines
Doktors der Medizin
genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. D. Neumeier

Prüfer der Dissertation: 1. Univ.-Prof. Dr. G. R. Thurmayer
2. Univ.-Prof. Dr. A. Neiß

Die Dissertation wurde am 01.06.2004 bei der Technischen Universität München
eingereicht und durch die Fakultät für Medizin am 22.07.2004 angenommen.

Inhaltsverzeichnis

	Seite
1. <u>Einleitung</u>	8
1.1. Grundlagen der Evidence-based Medicine	8
1.2. Szenario	10
1.3. Fragestellung	11
2. <u>Methoden</u>	12
2.1. Studienauswahl	12
2.1.1. Art der Studien.....	12
2.1.2. Themen der Studien	13
2.2. Fragebogen für die Überprüfung verschiedener Aspekte der einzelnen Studien	14
2.3. Das CONSORT- Statement	18
2.4. EBM- Bewertungsschema nach Sackett	22
2.4.1. Validität.....	24
2.4.1.1. Randomisierung.....	25
2.4.1.2. Berücksichtigung aller eingeschlossenen Patienten in der Gesamt- wertung	27
2.4.1.3. Intention-to-treat- Analyse	28
2.4.1.4. Verblindung	29
2.4.1.5. Gleichbehandlung	30
2.4.1.6. Ähnlichkeit der Patienten.....	31
2.4.2. Wichtigkeit.....	31
2.4.3. Anwendbarkeit	33
2.5. Erstellung eines objektiveren Bewertungssystems für die Validität von Therapie- studien	36
2.5.1. Randomisierung	37
2.5.2. Berücksichtigung aller eingeschlossenen Patienten in der Gesamtwertung	38
2.5.3. Intention-to-treat- Analyse	39
2.5.4. Verblindung	40
2.5.5. Gleichbehandlung	40

2.5.6. Ähnlichkeit der Patienten.....	41
2.5.7. Vorgehen bei der Zusammenfassung der Ergebnisse aus der objektiven Bewertung	42
2.6. Vorgehen bei der Erstellung eines unsystematischen Reviews zur Therapie der Multiplen Sklerose	43
3. <u>Ergebnisse</u>	45
3.1. Ergebnisse der Untersuchung von Validität, Wichtigkeit und Anwend- barkeit der Therapiestudien	45
3.1.1. Auswertung des Fragebogens für verschiedene Aspekte der verwendeten Studien	45
3.1.1.1. Ergebnisse der Datensammlung über die in die Studien einge- schlossenen Patienten und den Studienablauf	46
3.1.1.2. Ergebnisse der Untersuchung nach dem CONSORT- Statement	49
3.1.1.2.1. Ergebnisse der Untersuchung von Titel, Zusammenfassung und Einleitung	52
3.1.1.2.2. Ergebnisse der Untersuchung des Material- und Methoden- Teils.....	53
3.1.1.2.3. Ergebnisse der Untersuchung des Ergebnis- Teils	61
3.1.1.2.4. Ergebnisse der Untersuchung der Diskussion	69
3.1.2. Ergebnisse der Bewertung von Therapiestudien zur Multiplen Sklerose nach dem Sackett'schen EBM- Schema	74
3.1.2.1. Überprüfung der Validität mithilfe des objektiveren Bewertungs- systems	74
3.1.2.1.1. Randomisierung	74
3.1.2.1.2. Anzahl am Ende verbleibender Patienten.....	77
3.1.2.1.3. Intention-to-treat-Analyse	79
3.1.2.1.4. Verblindung	80
3.1.2.1.5. Gleichbehandlung.....	82
3.1.2.1.6. Ähnlichkeit der Behandlungsgruppen.....	84
3.1.2.1.7. Sonstige die Validität beeinflussende Faktoren	86
3.1.2.2. Überprüfung der Wichtigkeit	89

3.2. Review über Therapiestudien bei Multipler Sklerose	90
3.2.1. Therapieoptionen	90
3.2.1.1. Interferone	90
3.2.1.2. Cladribin.....	99
3.2.1.3. Monoklonale Antikörperbehandlung	100
3.2.1.4. TNF- Neutralisation	103
3.2.1.5. Bestrahlung in Kombination mit Kortikosteroiden	104
3.2.1.6. Sulfasalazin	105
3.2.1.7. AG284	107
3.2.1.8. Mitoxantron	108
3.2.1.9. Weitere Behandlungsmethoden der MS aus der weiterführenden Sekundärliteratur.....	110
3.2.1.9.1. Glatirameracetat	110
3.2.1.9.2. Azathioprin.....	112
3.2.1.9.3. Cyclophosphamid.....	114
3.2.1.9.4. Methotrexat	116
3.2.1.9.5. Cyclosporin A	117
3.2.1.9.6. Glukokortikosteroide	118
3.2.2. Bedeutung der Magnetresonanztomographie bei Multipler Sklerose	119
3.2.2.1. Korrelation von Magnetresonanztomographie-Ergebnissen mit dem Behinderungsgrad von MS-Patienten	119
3.2.2.2. Prädiktiver Wert der Magnetresonanztomographie bei Patienten mit einem ersten demyelinisierenden Ereignis	120
 4. Diskussion	 122
4.1. Verbesserung der Validitätsbeurteilung durch Verwendung eines objektivier- ten Bewertungssystems	122
4.2. Erkenntnisse durch die Validitätsbewertung der eingeschlossenen Therapie- studien und Nutzen der Evidence-based Medicine für die Durchführung einer wissenschaftlich gestützten objektiven Medizin	124
4.3. Vorschläge für die Verbesserung der Wichtigkeit und Anwendbarkeit von Therapiestudien	135
4.4. Vergleich der Therapiemöglichkeiten bei Multipler Sklerose	137
4.5. Stellenwert der Magnetresonanztomographie bei Multipler Sklerose.....	142

5. <u>Zusammenfassung</u>	144
6. <u>Literaturverzeichnis</u>	146
7. <u>Verzeichnis der Abbildungen, Tabellen und Diagramme</u>	158
7.1 Abbildungsverzeichnis	158
7.2 Tabellenverzeichnis	159
7.3 Diagrammverzeichnis	160
8. <u>Anhang</u>	161

Abkürzungsverzeichnis

AAA	Annual accumulation of lesion activity
ANOVA	Analysis of variance
Anteg	Antegren
ARR	Absolute Risikoreduktion
ARI	Absolute risk increase
Aza	Azathioprin
BMI	Body Mass Index
BMJ	British Medical Journal
BPF	Brain parenchymal fraction
CER	Control event rate
CI	Konfidenzintervall
Clad	Cladribin
CONSORT	Consolidated standards of reporting trials
Cop1	Copolymer 1 = Glatirameracetat
CPMS	Chronisch progressive Multiple Sklerose
CU	Combined unique
Cycl	Cyclophosphamid
EBM	Evidence-Based Medicine
EDSS	Expanded disability status scale
EER	Experimental event rate
ERMS	Exacerbating relapsing multiple sclerosis
ET	Experimentelle Therapie
FLAIR	Fluid attenuated inversion recovery
FU	Follow-up
Gd	Gadolinium
GHQ	General Health Questionnaire
GNDS	Guy's Neurological Disability Scale
GNRS	Gripps Neurologic Rating Scale
HD	High dose
HLA	Humanes Leukozytenantigen
i.d.R.	In der Regel
IFN	Interferon
IgG	Immunglobulin G
i.m.	Intramuskulär
IMSE	Institut für medizinische Statistik und Epidemiologie
ITT	Intention-to-treat- Analyse
i.v.	Intravenös
JAMA	Journal of the American Medical Association

KM	Kontrastmittel
LD	Low dose
Len	Lenercept
LK	Lymphknoten
MBP	Myelinbasenprotein
MeSH	Medical Subject Headings
MIU	Million international units
MR	Magnetresonanz-
MRI	Magnetresonanz-Imaging
MRT	Magnetresonanztomograph/-ie
MS	Multiple Sklerose
NEJM	New England Journal of Medicine
nIFN	Natürliches Interferon
NK-Zellen	Natürliche Killerzellen
NNT	Number needed to be treated
NRS	Scripps Neurologic Rating Scale (sh.a. SNRS)
n.sign./ n.s.	Nicht signifikant
NW	Nebenwirkung/-en
OR	Odd's ratio
pAVK	Periphere arterielle Verschlusskrankheit
PD	Proton density
PEER	Patient expected event rate
PL	Placebo
PMS	Progressive Multiple Sklerose
RCT	Randomised controlled trial
RMS	Relapsing multiple sclerosis
RPMS	Relapsing-progressive multiple sclerosis
RRI	Relative Risikozunahme
RRMS	Relapsing-remitting multiple sclerosis
RRR	Relative Risikoreduktion
s.c.	Subkutan
SF	Standardfehler
SLCMSR	Sylvia Lawry Centre for Multiple Sclerosis Research
SNRS	Scripps Neurologic Rating Scale
SPMS	Sekundär progressive Multiple Sklerose
S-TLI	Sham total lymphoid irradiation
Sulf	Sulfasalazin
TLI	Total lymphoid irradiation
TNF	Tumornekrosefaktor
UAW	Unerwünschte Arzneimittelwirkung/-en
z.T.	Zum Teil

1. Einleitung

1.1. Grundlagen der Evidence-based Medicine

Die philosophischen Wurzeln der Evidence-based Medicine (EBM) gehen bereits in die Mitte des 19. Jahrhunderts zurück. Damals war Paris das Zentrum von Überlegungen, dass praktizierte Medizin auf gesicherten Erkenntnissen beruhen müsse (Sackett 1997, S.644 [136]). Als wissenschaftliche Beweise wurden jedoch zu diesem Zeitpunkt Ergebnisse aus Expertenmeinungen und Einzelfalluntersuchungen betrachtet, und die randomisierte, kontrollierte Studie blieb bis in die 60er Jahre des 20. Jahrhunderts eine Seltenheit (Users' Guides 1992, S.2420 [154]). Der Glaube an die Autorität einzelner Gelehrter brachte sogar Kuriositäten hervor wie die Behauptung von Aristoteles, die Frau habe weniger Zähne als der Mann. Ohne Überprüfung wurde dieses Postulat als Faktum angesehen und hatte bis über das Mittelalter hinaus Bestand (Köbberling 1997 [180]). Ein anderes Beispiel für das Vertrauen in das „Wissen“ von Koryphäen ist die Aussage von Galen, eine Wunde müsse eitern, um heilen zu können, weswegen Tausende von Menschen im Mittelalter den Tod fanden.

Schon 1932 veröffentlichte Paul Martini seine „Methodenlehre der therapeutischen Intervention“ (Sackett 1999, S.VI [134]), doch erst ab etwa 1965 vollzog sich in der Medizin langsam ein Paradigmenwechsel, der schließlich dazu führte, dass heute die randomisierte kontrollierte Studie als die beste Möglichkeit anerkannt ist, die Wirksamkeit eines Medikaments zu überprüfen. Dies ist unter anderem das Verdienst von Archie Lemman Cochrane (siehe Anhang, Abbildung 1), der 1972 sein revolutionäres Buch „Effectiveness and Efficiency. Random Reflections on Health Services“ verfasste (Cochrane 1972 [68]). Grund für seine Überlegungen war der Wille, ausgehend von immerwährender Knappheit der Ressourcen, die zur Verfügung stehenden Mittel für eine optimale Patientenversorgung einzusetzen. Cochrane war der Erste, der dafür eindeutig die außerordentliche Wichtigkeit randomisierter kontrollierter Studien darlegte. Aufgrund seiner Erkenntnisse wurde 1987 erstmals ein systematisches Review über Betreuung während Schwangerschaft und Kindheit durchgeführt, das er als „ein(en) echte(n) Meilenstein in der Geschichte von randomisierten Studien und in der Entwicklung der Patientenversorgung“ (www.cochrane.org [166]) bezeichnete. Sein Engagement und die Unterstützung anderer führten schließlich 1993, fünf Jahre nach seinem Tod zur Gründung der Cochrane Collaboration. Mithilfe dieser Institution werden immer mehr randomisierte kontrollierte Studien in systematischen Reviews und Metaanalysen zusammengefasst und der medizinischen Öffentlichkeit zugänglich gemacht. So enthielt zum Beispiel Ausgabe 3/2003 der Cochrane-

ne Library 85 neue bei mittlerweile insgesamt 1754 durch die Cochrane Collaboration durchgeführten systematischen Reviews (www.cochrane.org [166]).

Schon 1990 jedoch prägte Gordon Guyatt den Begriff der „evidence based medicine“ (www.fhs.mcmaster.ca [170]). Kurze Zeit später gründete er die „Arbeitsgruppe Evidence-based Medicine“, an der unter anderem David Sackett, Scott Richardson, William Rosenberg und Brian Haynes, die späteren Herausgeber des Buches „Evidence-based Medicine – How to Practice & Teach EBM“ (Sackett 1996 [133]), teilnahmen. Es ist das Verdienst von David Sackett, dass für eine bis dahin als mehr oder weniger selbstverständlich aufgefasste Vorgehensweise ein konkretes Konzept erarbeitet wurde, „(...) bei dem sich die Beweise auf transparente reproduzierbare Daten beziehen, deren Glaubwürdigkeit an den strengen Methoden der klinischen Epidemiologie gemessen werden“ (Sackett 1999, S.VII [134]). Ein weiteres Verdienst von David Sackett ist, dass die Methoden der Evidence-based Medicine der medizinischen Öffentlichkeit zugänglich gemacht wurden und somit als Hilfestellung bei der Diagnostik und Therapie dem klinisch tätigen Arzt zur Verfügung stehen. Der sechsteilige Fragebogen für die Bewertung klinischer Studien ist heute die am weitesten verbreitete Form der Umsetzung der Evidence-based Medicine (Sackett 1999 [134]). Sacketts Definition für Evidence-based Medicine lautet wie folgt: „EBM ist der gewissenhafte, ausdrückliche und vernünftige Gebrauch der gegenwärtig besten externen, wissenschaftlichen Evidenz für Entscheidungen in der medizinischen Versorgung individueller Patienten. Die Praxis der EBM bedeutet die Integration individueller klinischer Expertise mit der bestmöglichen externen Evidenz aus systematischer Forschung“ (Sackett 1997, S.644 [136]). Heute geht man sogar soweit, auch diagnostische Tests, prognostische Aussagen oder chirurgische Therapieverfahren mithilfe randomisierter kontrollierter Studien zu untersuchen. Evidence-based Medicine sorgt für Entscheidungen, die auf klinisch relevanten Studien basieren. Studien basieren

Umsetzung der EBM in 5 + 1 Schritten
1. Formulierung einer beantwortbaren Frage
Zusatzschritt: Beantwortung der Frage mit interner Evidence
2. Suche nach externer Evidence
3. Bewertung der gefundenen externen Evidence
4. Integration der externen Evidence in die bestehende interne Evidence, Synthesebildung und Anwendung
5. Bewertung des erzielten Ergebnisses

Tabelle 1: Umsetzung der EBM in 5 + 1 Schritten nach Sackett 1997, S.XV [136] und Porzsolt 2001 [125].

und minimiert Vorgehensweisen, die aufgrund von überholten Lehrbüchern, lokalen, autoritätsgestützten Verhaltensmustern, Broschüren für Produktmarketing oder subjektiven Expertenmeinungen beruhen.

Mittlerweile existiert eine Fülle von Artikeln, in denen Ärzte angewiesen werden, wie medizinische Literatur

gesucht, evaluiert und interpretiert werden kann. Medizinische Journals haben ihr Artikel-Layout dahingehend geändert, dass der Methodenteil sofort ins Auge fällt, und es existieren neue Formen von Fachgesellschaften und Journals, in denen Publikationen auf ihre

Validität und klinische Relevanz untersucht werden. Leitlinien als direkte Zusammenfassung und Umsetzung wichtiger und valider Studien haben ihren festen Platz in der Patientenversorgung (Users´ Guides 1992, S.2421 [154]).

Um EBM zu praktizieren, ist es am günstigsten, in den sogenannten „5 + 1 Schritten“ vorzugehen, die in Tabelle 1 vorgestellt werden. Die fünf ursprünglichen Schritte wurden bereits von Sackett und Kollegen (Sackett 1999, S.XV [134]) aufgestellt, während der Zusatzschritt nachträglich von Porzsolt und Kollegen (Porzsolt 2000 [124]) eingefügt wurde, um der klinischen Erfahrung behandelnder Ärzte Rechnung zu tragen. In Schritt 1 geht es um die Formulierung einer vierteiligen Frage, die zunächst den Patienten und sein Problem enthalten sollte, dann die Methode zur Lösung des Problems sowie die alternative Handlungsmöglichkeit und schließlich den Zielparameter. Der für die vorliegende Arbeit wichtigste Schritt ist Nummer 3 mit der Bewertung von Validität, Wichtigkeit und Anwendbarkeit der eingeschlossenen Therapiestudien.

Bei der Übersetzung des Wortes Evidence gibt es in Deutschland immer noch Schwierigkeiten, da die Bedeutung im Englischen – das Vorhandensein von Beweisen für oder gegen eine These – nicht mit dem Begriff Evidenz im Deutschen, der eine „unmittelbare und vollständige Einsichtigkeit, Deutlichkeit, Gewißheit“ (Sackett 1999, S.V [134]) bedeutet, übereinstimmt. Da aber im Begriff „Evidence-based Medicine“ eher der Sinn des englischen Wortes gemeint ist, wird in der vorliegenden Arbeit stets der englische Begriff verwendet.

1.2. Szenario

Grundlage jeder Umsetzung der Evidence-based Medicine in der Praxis ist ein konkreter Fall eines Patienten, für dessen Problem relevante und passende medizinische Literatur gesucht, bewertet und umgesetzt werden muss. Dies spiegelt sich in Schritt 1 aus Tabelle 1 wieder.

Exemplarisch wird für die vorliegende Arbeit folgendes Szenario aufgestellt: Ein 39-jähriger männlicher Patient mit schubweiser Multipler Sklerose (MS) und ohne weitere relevante Vorerkrankungen kommt in die Neurologie eines Universitätskrankenhauses, weil er gerade einen erneuten Schub der Multiplen Sklerose erleidet. Sein Wert auf der Expanded-Disability-Status-Scale (EDSS) (Kurtzke 1983 [97]) beträgt 3,0 bei einer bisherigen Krankheitsdauer von 4,5 Jahren. Der Patient fragt den behandelnden Arzt, welche die derzeit beste Therapie für seine Erkrankung ist, woraufhin der Arzt eine Medline-Suche durchführt und die gefundenen Studien nach den Kriterien der Evidence-based

Medicine bewertet, um die für den Patienten passenden, aussagekräftigen und validen Studien herauszufinden.

1.3. Fragestellung

Die vorliegende Arbeit befasst sich mit der Bewertung von Studien, die unterschiedliche Therapiemöglichkeiten bei Multipler Sklerose untersuchen und für die Evaluation des Behandlungserfolgs unter anderem die Magnetresonanztomographie (MRT) verwendeten. Dabei wurde die Multiple Sklerose als Grunderkrankung gewählt, weil mit dem Sylvia Lawry Centre for Multiple Sclerosis Research (SLCMSR) seit 2001 im Zusammenhang mit dem Institut für medizinische Statistik und Epidemiologie der technischen Universität München (IMSE) erstmalig ein Zentrum entstanden ist, in dem Ergebnisse weltweiter Studien über die MS archiviert werden und so für zukünftige Untersuchungen zu dieser Krankheit zum Beispiel virtuelle Placebogruppen errechnet werden können. Es wurde besonderes Augenmerk auf die Diagnosemöglichkeit MRT gelegt, da dies eine der wenigen Möglichkeiten ist, den Therapieerfolg bei der Behandlung der MS mittels Bildgebung zu objektivieren, andererseits aber gerade dieses diagnostische Mittel bei MS noch umstritten ist.

Ziel dieser Arbeit ist es zunächst, mit Hilfe des CONSORT-Statements (siehe Abschnitt 2.3.) als Voraussetzung für die anschließende Auswertung nach EBM-Kriterien (siehe Abschnitt 2.4.) positive Aspekte in der Durchführung von Therapiestudien und deren Berichterstattung hervorzuheben sowie Mängel aufzudecken. Daraus soll einerseits dargestellt werden, welche Aspekte von CONSORT und EBM bereits häufig erfüllt sind, und andererseits sollen Verbesserungsvorschläge für zukünftige Therapiestudien entwickelt werden. Anhand der Resultate der Untersuchungen wird der von Sackett und Kollegen entworfene Fragebogen (Sackett 1999 [134]) in dieser Arbeit erweitert und objektiviert (siehe Abschnitt 2.5.).

Des Weiteren folgt im zweiten Teil der Arbeit ein kurzes Review über die untersuchten und als voll oder eingeschränkt valide eingestuft Studien. Dabei wird über die Wirkungen der verschiedenen Therapiemethoden auf klinische Parameter und MRT-Veränderungen der in den Studien eingeschlossenen Patienten berichtet, und es wird diskutiert, ob die MRT ein geeignetes Mittel für die Überprüfung des Therapieerfolgs darstellt.

2. Methoden

2.1. Studienauswahl

In diese Arbeit einzuschließende Studien wurden mithilfe Medline-Suche (www.ncbi.nlm.nih.gov/PubMed/ [175]) über den MeSH-Browser von PubMed erfasst. MeSH steht für „Medical Subject Headings“ und subsumiert eine große Anzahl medizinischer Begriffe, die im MeSH-Browser über Verknüpfungsbäume miteinander verbunden sind. Der gesuchte Fachausdruck wird dabei immer im Kontext mit anderen Begriffen gesehen.

Suchkriterium für die vorliegende Arbeit war „Multiple Sclerosis“ in Verbindung mit „Magnetic Resonance Imaging“. Dazu wurde die von Medline zur Verfügung gestellte sogenannte „History-Funktion“ genutzt. Diese ermöglicht es, zunächst jeden Begriff einzeln zu suchen und die Resultate anschließend nach Anzeige der Ergebnisse miteinander zu verknüpfen. Medline erstellt dabei automatisch eine Verlaufsliste, die man unter „History“ findet. Um nun aus zwei Resultaten die Studien zu finden, die beide gesuchte Begriffe enthalten, gibt man die Nummern der entsprechenden Suchergebnisse, verbunden mit einem „#“ und verknüpft mit einem groß geschriebenen „AND“, ein. Angenommen, man möchte diejenigen Studien anzeigen lassen, die sowohl Resultat 1 der Suche als auch Resultat 2 enthalten, so gibt man in der Suchzeile „#1 AND #2“ ein. Als Ergebnis erhält man Resultat 3, die Schnittmenge aus Nummer eins und Nummer zwei.

Zusätzlich besteht weiterhin die Möglichkeit, unter „Limits“ die Suchergebnisse einzuschränken, zum Beispiel auf nur englischsprachige Studien, auf bestimmte Daten, auf ein bestimmtes Geschlecht und vieles andere.

2.1.1. Art der Studien

Die Literatursuche für die vorliegende Arbeit war limitiert auf randomisierte kontrollierte Therapiestudien, veröffentlicht zwischen dem 1. Januar 1997 und dem 31. Dezember 2000, und wurde durchgeführt am 7. August 2001. Von den 46 per MeSH-Abfrage identifizierten Schriften erfüllten 38 die genannten Kriterien, fünf Studien hatten trotz vorheriger Einschränkung der Suche Beurteilung von Diagnosemöglichkeiten zum Ziel (Filippi, J Neurol Sci 1998 [6]; Brunetti 1997 [1]; Molyneux 2000 [21]; Fisher 2000 [9]; Molyneux 1998 [23]), eine befasste sich mit Infektionen des oberen Respirationstraktes bei Multipler Sklerose (Edwards 1998 [5]), eine beschrieb die Untersuchung der Häufigkeit einer ICAM-1 Exon 6 Genlokus Mutation vom Typ K469 (Mycko 1998 [24]) und eine war nur der Vorbericht zu noch laufenden Untersuchungen (Hommes 2000 [11]). Nach der Suche wurden die Studien in einer Accessdatenbank erfasst und sortiert nach Autoren, Titel, Journal, Erscheinungsdatum- und Band.

Außerdem wurde zu diesem Zeitpunkt bereits notiert, ob in den Publikationen ein informed consent erwähnt war, wie viele Patienten mit welcher MS-Form eingeschlossen worden waren, wie viele Ärzte involviert waren und welche Studienform vorlag (z.B. Multicenter-Studie). Weiter wurde festgehalten, ob die Patienten nach „Intention-to-treat“ ausgewertet worden waren, ob eine Table 1 enthalten und die Vergleichsgruppen ähnlich waren, wie randomisiert worden war, ob die Patienten gleich behandelt worden waren, wie lange die Nachbeobachtung gedauert hatte und ob systemische Fehler vorgelegen hatten. Grundsätzlich positive und negative Aspekte der einzelnen Studien wurden notiert.

2.1.2. Themen der Studien

Von den bewerteten Studien befassten sich acht mit Interferon β -1a (The Once Weekly Interferon for MS Study Group 1999 [42]; Rudick 1999 [34]; Simon 1998 [36]; Li 1999 [18]; Jacobs 2000 [13]; Jacobs 2000 [12]; Paolillo 1999 [30]; Koudriavtseva 1998 [17]), vier mit Interferon β -1b (Zhao 2000 [46]; Miller 1999 [20]; Molyneux 2000 [22]; Kastrukoff 1999 [15]), eine mit einem Vergleich von Interferon β -1a und Interferon β -1b (Koch-Henriksen 2000 [16]),

Experimentelle Therapie	Untersucht in Studie (Nummer entsprechend Literaturverzeichnis)
Interferone:	
IFN β -1a	12, 13, 17, 18, 30, 34, 36, 42
IFN β -1b	15, 20, 22, 46
IFN β -1a vs. IFN β -1b	16
Natürliches IFN β	31
IFN a-2a	25, 26
Monoklonale Antikörpertherapie	2, 37, 38, 40, 43, 44
Cladribin	7, 8, 32
Methylprednisolon	28, 29
Bestrahlung + Steroide	3, 33
Linomid	39, 45
Mitoxantron	4, 19
Myelin basic proteins	10, 14
Sulfasalazin	27
Lenercept	41
Amantadin	35

Tabelle 2: Untersuchte Studien und deren Themen.

eine mit natürlichem Interferon beta (Patti 1999 [31]) und zwei mit Interferon a-2a (Nortvedt 1999 [26]; Myhr 1999 [25]). Des weiteren behandelten sechs Studien monoklonale Antikörper-Therapie (Tubridy 1999 [43]; Coles 1999 [2]; Sørensen 1998 [37]; Sørensen 2000 [38]; Teksam 2000 [40]; van Oosten 1997 [44]), drei Cladribin (Filippi 2000 [7]; Filippi 2000 [8]; Rice 2000 [32]), zwei Methylprednisolon (Optic Neuritis Study

Group 1997 [29]; Oliveri 1998 [28]), zwei eine Kombination aus Bestrahlung und Kortikosteroiden (Cook 1997 [3]; Rohowsky-Kochan 1997 [33]), zwei Linomid (Wolinsky 2000 [45]; Tan 2000 [39]), zwei Mitoxantron (Millefiorini 1997 [19], Edan 1997 [4]), zwei myelin basic proteins (gelöstes DR2:MBP84-102(AG284): Goodkin 2000 [10]; modifizierte Peptidliganden:

Kappos 2000 [14]) und jeweils eine Sulfasalazin (Noseworthy 1998 [27]), Lenercept (The Lenercept Multiple Sclerosis Study Group 1999 [41]) und Amantadin (Sailer 2000 [35]). Tabelle 2 gibt einen Überblick über die Themen der Studien und die korrespondierende Nummer im Literaturverzeichnis.

2.2. Fragebogen für die Überprüfung verschiedener Aspekte der einzelnen Studien

Validität, Wichtigkeit und Anwendbarkeit sind die drei klassischen Kriterien der EBM. Voraussetzung für eine Bewertung dieser Merkmale ist jedoch eine hohe Qualität der Durchführung und des Berichts über randomisierte, kontrollierte Studien (randomised controlled trials = RCTs). Untersuchungen haben ergeben, dass unzureichende Berichterstattung und inadäquates Studiendesign von RCTs zu verfälschten Studienergebnissen führen (Altman 2001, S.663 [48]). Diese Verfälschung zeigt sich in systematischen Fehlern, also nicht durch den Zufall bedingten Streuungen um das unbekannt wirkliche Ergebnis (Perleth, S.101 [121]). Um derartige Fehler zu verhindern, wurde von Altman et al. in der CONSORT-Gruppe das CONSORT-Statement erarbeitet, das mittlerweile von vielen Fachzeitschriften, zum Beispiel dem British Medical Journal (Altman 1996 [50]), dem Canadian Medical Association Journal (Huston 1996 [91]), dem Journal of the American Medical Association (Meinert 1998 [101]), der Atención Primaria (Ausejo 1998 [59]), der Nederlands Tijdschrift Voor Geneeskunde (Ned Tijdschr Geneesk 1998 [116]), den Annals of Internal Medicine (Moher 2001 [110]) und anderen anerkannt und adaptiert wurde und auf das im Gliederungspunkt 2.3. dieser Arbeit noch genauer eingegangen wird. Neben dem CONSORT-Statement mit der Frage nach der Qualität der Berichterstattung als Voraussetzung für eine Bewertung nach EBM gibt es aber auch weitere Ansätze zur Überprüfung der Güte in RCTs. Drei sollen hier exemplarisch für mögliche Vorgehensweisen bei der Qualitätskontrolle genannt werden. So erstellten Schäfer et al. 1999 die „Empfehlungen für die Erstellung von Studienprotokollen (Studienplänen) für klinische Studien“ (Schäfer 1999 [138]), der Weltärztebund 1964 die bis 1996 viermal revidierte „Deklaration von Helsinki“ (Weltärztebund 1996 [188]) und die University of Oxford das DISCERN-Instrument (Abteilung Epidemiologie 2002: „Allgemeine Anleitung“ und „Das DISCERN-Instrument“ [162]). Die Ansätze dieser Bewertungsmöglichkeiten setzen ihren Schwerpunkt an anderer Stelle als das CONSORT-Statement, beschreiben aber ebenfalls wichtige Kriterien. Schäfer et al. fordern vor allem einen konkreten Studienplan und dessen exakte Einhaltung. Sie schreiben: „Als Grundlage für die Durchführung biomedizinischer Forschung am Menschen sollte stets ein detaillierter Studienplan (...) vorliegen, der die Voraussetzung, die konkrete Fragestellung (...), das Studiendesign, die Durchführung und die Organisation und Auswertung des Vorhabens beschreibt sowie Anhaltspunkte für die Interpretation verschiedener möglicher Studienergebnisse gibt. (Er) (...) dient (...) als verbindliche

Festlegung des Studienvorgehens in Durchführung, Auswertung und Ergebnisinterpretation (...). Er dient Gutachtergremien (...) als Unterlage zur Beurteilung der Studie.“ (Schäfer 1999, S.142 [138]). Der Weltärztebund hingegen empfiehlt eine verantwortungsvolle Zusammenarbeit zwischen Arzt und Patient bei der Durchführung von klinischen Studien, um „(...) die wissenschaftliche Kenntnis zu fördern und der leidenden Menschheit zu helfen (...)“ (Weltärztebund, S.1 [188]). Es wird gefordert, dass „(...) der Arzt die Freiheit haben (muss,) neue diagnostische und therapeutische Maßnahmen anzuwenden (...)“, aber dass er die „(...) möglichen Vorteile, Risiken und Störungen des Befindens (...) gegen die Vorzüge der bisher bestehenden (...) Methoden (...)“ abwägen muss (Weltärztebund, S.3 [188]). Das DISCERN-Instrument schließlich schlägt ein Bewertungsschema aus 16 Fragen in drei Abschnitten vor. Abschnitt I fragt nach der Zuverlässigkeit einer Studie, Abschnitt II nach der Qualität der Informationen zu den Behandlungsalternativen und Abschnitt III nach der Gesamtbewertung einer Publikation. Dabei sind die Fragen im Abschnitt I noch am ehesten mit den CONSORT- und den EBM-Fragen zu vergleichen, indem zum Beispiel evaluiert wird, ob die Ziele der Publikation eindeutig sind (Frage 1), ob klare Angaben zu den Informationsquellen für eine Publikation vorliegen (Frage 4) oder ob die Publikation ausgewogen und unbeeinflusst geschrieben wurde (Frage 6). Dabei geht das DISCERN-Instrument allerdings nicht sehr in die Tiefe, die Bewertung bleibt wesentlich subjektiver und oberflächlicher, und es werden bei weitem nicht so viele Bereiche wie durch das CONSORT-Statement abgedeckt. Während das DISCERN-Instrument also einen Versuch in Richtung systematische Bewertung einer Studie liefert, betrachten Schäfer et al. die Einhaltung des Studienplans als Qualitätskriterium für klinische Studien, und der Weltärztebund legt auf die Eigenverantwortlichkeit des Studienleiters bei der Überprüfung der Nützlichkeit einer Studie in Zusammenarbeit mit dem Patienten Wert. Alle drei Forderungen gehen jedoch bei weitem nicht so systematisch vor und sind nicht so breit angelegt wie das EBM-Schema nach Sackett, das deshalb als das derzeit beste und praktikabelste Schema zur Überprüfung von Validität, Wichtigkeit und Anwendbarkeit gelten muss. Damit eine Arbeit mit diesem Schema effektiv und objektiv funktionieren kann, braucht es eine fundierte Berichterstattung, für die das CONSORT-Statement eine sehr gute Grundlage bietet. Das Hauptanliegen des CONSORT-Statements ist die Bereitstellung einer Checkliste und eines Flussdiagramms, die bei der Berichterstattung von RCTs verwendet werden sollen (Altman 2001, S.663 [48]). Selbstverständlich sind im CONSORT-Statement auch die Zuverlässigkeit einer Publikation, die Einhaltung eines Studienplans und die Verantwortung des Arztes wichtige Kriterien. Damit sind die Forderungen des Weltärztebundes, der University of Oxford und von Schäfer und Kollegen auch Teil des CONSORT-Statements. Dieser Ansatz ist derzeit der vielversprechendste in Hinblick auf die Vermeidung von Bias bei der Durchführung eines RCTs.

Aus diesem Grund wurde für die vorliegende Arbeit eine Checkliste erarbeitet, die in Anlehnung an das CONSORT-Statement die für MRT-gestützte MS-Studien wichtigen Aspekte mit den für alle Therapiestudien geltenden Forderungen kombiniert (Abbildung 2).

<u>Untersuchte Studie:</u>	
Zeitschrift	
Autoren	
Jahr, Nummer, Seiten	
Artikel	
Betrifft Arbeitsthese	
<u>Fragen:</u>	
Liegt eine Diagnostik (=D)-, Prognose (=P)- oder Therapiestudie (=T) vor?	
<u>Zu den Patienten:</u>	
Wie viele Patienten mit welchem Geschlecht nahmen an der Studie teil?	, davon ?: ?:
Wie alt waren die Patienten?	
Gab es physische Untersuchungen der Patienten?	
In welchem Stadium der Erkrankung befanden sich die Patienten (akut/chron.)?	-
EDSS-Score?	-
Wie lange bestand die Erkrankung schon?	
Änderung des Scores während der Studie?	
Welchen Typ der MS hatten die Patienten?	
Welche Behandlung erhielten die teilnehmenden Patienten während der Studie?	-
-Wurde sie strikt eingehalten?	-
-Nebenwirkungen der Behandlung?	-
<u>Zum Studienablauf:</u>	
Wie war das <u>Studienschema</u> aufgebaut? Wie häufig waren: -die Medikamentengabe, -MRI-Scans, -Vergleichsmethoden? -...	- - - -
Wenn KM, dann welche? -Dosierung?	- -
Wie viele Auswerter gab es?	
Wie lange war das Follow-up?	
Wurde die Behinderung der Patienten gemessen? Wenn ja, wie?	EDSS ? SNRS ? ... ? ... ?

Abbildung 2: Checkliste für MRT-gestützte MS-Therapiestudien inklusive Forderungen aus dem CONSORT-Statement (fortgesetzt auf der folgenden Seite).

Zum Artikel:

<u>Abstract:</u> Genannt?:	
-Ziel der Studie?	
-Hintergrund?	
-Methoden?	
-Resultate?	
-Schlussfolgerung?	
-Anzahl der Patienten?	
Befinden sich Literaturangaben (Zitation) im Text?	
Stimmen die Angaben aus dem Text mit den gezeigten Tabellen und Abbildungen überein?	
<u>Einführung:</u>	
Wissenschaftlicher Hintergrund + Erklärung der Gründe für Studie?	
<u>Material und Methoden:</u>	
Genauere Beschreibung der <u>Patienten</u> ?	-
Auswahlkriterien; Rahmenbedingungen und Orte der Patientendatenerhebung	- -
Alter, Geschlecht	- -
Krankheitsdauer, EDSS	- -
Rasse (weiß, schwarz o.ä.)	-
Anzahl der Exazerbationen (in 2 Jahren vor Studienbeginn)	-
Wie Patientenzahl bestimmt?	-
Erklärung von Zwischenanalysen und Abbruchkriterien	-
Methode der Randomisation beschrieben? (Wie Sequenz hergestellt, wie geheimgehalten, wie durchgeführt?)	
Spezifische Ziele u. Hypothesen der Studie?	
Genauere Beschreibung der <u>Therapiemethoden</u> (wie und wann)...	
... und der <u>Untersuchungstechniken</u> :	-
Dicke und Anzahl der Schichten,	-
Auflösung der Images,	-
Vorgehensweise bei Anfertigung der Images (Lagerung der Pat., Reihenfolge der Einzelschichten, Ausschaltung von Fehlern, ...),	-
Sonstiges	-
Genauere Beschreibung der <u>Auswertung und Analyse</u> ?:	-
Training der Untersucher,	-
Statistische Methoden (welche für was?), Bewertung der Läsionen nach Größe und Lage,	-
Beschreibung, wann als Läsion gezählt und wann nicht,	-
Sonstiges	-
<u>Resultate:</u>	
Nebenwirkungen der Medikamente/ Kontrastmittel?	- -
Ausführliches Flussdiagramm?	-
Angaben über klinische Daten?:	-
Alter, Krankheitsdauer, EDSS, Sonstige	- - - -
Daten/ Zeitpunkte d. Rekrutierung und des Follow-up?	-
Genauere Angaben für jedes primary und secondary outcome?	-
Zusätzliche Untersuchungen (subgroup analyses), vorher im Studienprotokoll oder zusätzlich?	-
Angaben über MRI-Resultate?	-
Angaben der p-Werte? CI? Power?	- - -
Angaben der ARR, RRR, NNT, OR?	- - - -
<u>Diskussion:</u>	
Vor- und Nachteile der verglichenen Methoden beschrieben?	-
Ausblick in die Zukunft/ Verallgemeinerungsmöglichkeiten der Ergebnisse?	-
Erklärung der Ergebnisse im Kontext mit bestehender Evidenz (inkl. Quellen möglichen Bias, Impräzision etc.)	-

Abbildung 2: Fortsetzung Checkliste für MRT-gestützte MS-Therapiestudien inklusive Forderungen aus dem CONSORT-Statement

Für die Anfertigung dieses Fragebogens wurden zunächst die für MS relevanten Aspekte identifiziert, wie zum Beispiel der EDSS-Score der Patienten zu Beginn der Erkrankung, die Anzahl der Exazerbationen, der MS-Typ oder verschiedene MRT-Untersuchungstechniken. Anschließend wurden die Ergebnisse mit den wichtigen Fragen aus dem CONSORT-Statement (Altman 2001, S.665 [48]) integriert und daraus die oben stehende Checkliste erstellt.

Bei der Durchführung der Bewertung wurden die eingeschlossenen Studien anhand dieses Fragebogens untersucht, die Ergebnisse in einer Microsoft Access-Datenbank zwischengespeichert und anschließend die Publikationen mithilfe des Bewertungsschemas nach Sackett (Abbildung 5) mit EBM-Kriterien evaluiert.

2.3. Das CONSORT-Statement

Schon zu Beginn der 80er Jahre des letzten Jahrhunderts schrieben DerSimonian und Kollegen, dass Redakteure die Qualität von RCTs verbessern könnten, indem sie Autoren eine Liste mit erforderlichen Kriterien für eine Veröffentlichung an die Hand gäben (DerSimonian 1982, S.1236 [69]). Es dauerte jedoch noch bis zum Beginn der 90er Jahre, bis zwei Gruppen unabhängig voneinander Vorschläge für den Bericht über RCTs machten - die „Standards of Reporting Trials Group“ (The Standards of Reporting Trials Group 1994 [153]) und die „Working Group on Recommendations for Reporting of Clinical Trials in the Biomedical Literature“ (Working Group on Recommendations 1994 [159]). Aus diesen zwei Gruppen bildete sich schließlich die CONSORT-Gruppe. Dabei zeigt sich schon im Namen, zu welchem Zweck sich die Arbeitsgruppe formierte: CONSORT steht für „*Consolidated Standards of Reporting Trials*“ (Altman 2001, S.664 [48]), was bedeutet, dass fundierte Standards erarbeitet wurden, mit deren Hilfe man die Publikation von Studien transparenter gestalten kann. Bias soll vermieden werden, indem Wissenschaftler die von der CONSORT-Gruppe erarbeitete Checkliste (Abbildung 3) vor, während und nach der Durchführung einer randomisierten kontrollierten Studie anwenden (Falck-Ytter 2002 [165]). Denn: „Only high-quality research, in which proper attention has been given to design, will consistently eliminate bias“. (Altman 2001, S.686 [48]) Allerdings wird ausdrücklich darauf hingewiesen, dass das CONSORT-Statement nicht als „(...) quality assessment instrument“ (Altman 2001, S.664 [48]) verwendet werden soll. Dennoch wird mit dem CONSORT-Statement auch ein großer Schritt in Richtung Qualitätsverbesserung getan. Denn ein Autor, der weiß, worauf es in einer Studie ankommt, wird auch darauf achten, die für eine RCT wichtigen Kriterien zu erfüllen, weil er anderenfalls aufgrund der CONSORT-Checkliste berichten müsste, welche Aspekte bei der Studiendurchführung nicht berücksichtigt wurden. Im Gegensatz zu dem EBM-Schema nach

Sackett verlangt das CONSORT-Statement, ohne dabei eine Wertung vorzunehmen, nur nach einem exakten Bericht, wie etwas in einer Studie gemacht wurde.

Publikationsabschnitt	Beschreibung	Erwähnt auf Seite
Titel und Zusammenfassung	1 Zuordnung zu Therapiegruppen (z.B. „randomisierte Verteilung“, „randomisiert“ oder „randomisierte Zuweisung“)	
Einleitung		
Hintergrund	2 Wissenschaftlicher Hintergrund und Begründung der Studie.	
Methoden		
Probanden/Patienten	3 Einschlusskriterien der Probanden/Patienten, Studienorganisation und Ort der Studiendurchführung (z.B. im Krankenhaus oder nicht-stationär).	
Intervention/Behandlung	4 Präzise Angaben zu den geplanten Interventionen jeder Gruppe und zur Durchführung.	
Ziele	5 Genaue Ziele, Fragestellung und Hypothesen.	
Ergebnisse	6 Klar definierte primäre und sekundäre Zielkriterien und, gegebenenfalls, alle zur Optimierung der Ergebnisqualität verwendeten Methoden (z.B. Mehrfachbeobachtungen, Training der Prüfer).	
Fallzahlbestimmung	7 Wie wurden die Fallzahlen bestimmt und, falls notwendig, Beschreibung von Zwischenanalysen und Kriterien für einen vorzeitigen Studienabbruch.	
Randomisierung		
Erzeugung der Behandlungsgruppe	8 Methode zur Generierung der zufälligen Zuteilung, einschließlich aller Einzelheiten (wie z.B. Block-Randomisierung, Stratifizierung).	
Geheimhaltung der Behandlungsfolge (allocation concealment)	9 Durchführung der Zuteilung (z.B. numerierte Behälter; zentrale Randomisierung per Fax/Telefon), Angabe, ob Geheimhaltung bis zur Zuteilung gewährleistet war.	
Durchführung	10 Wer führte die Zuteilung durch, wer nahm die Probanden/Patienten in die Studie auf und wer teilte die Probanden/Patienten den Gruppen zu.	
Verblindung	11 Waren a) die Probanden/Patienten und/oder b) diejenigen, die die Intervention/Behandlung durchführten und/oder c) diejenigen, die die Zielgrößen beurteilten verblindet oder nicht verblindet. Wir wurde der Erfolg der Verblindung evaluiert?	
Statistische Methoden	12 Statistische Methoden zur Bewertung der primären Zielkriteriums; weitere Analysen, wie z.B. Subgruppenanalysen und adjustierte Analysen.	
Ergebnisse		
Ein- und Ausschlüsse	13 Anzahl der Studienteilnehmer für jede Behandlungsgruppe, die a) gemäß Randomisierung gebildet wurde, b) tatsächlich die geplante Behandlung/Intervention erhalten haben, c) die Studie protokollgemäß beendeten, d) in der Analyse des primären Zielkriteriums berücksichtigt wurden (Darstellung in Flussdiagramm empfohlen), Beschreibung von Protokollabweichungen mit Angabe von Gründen.	
Aufnahme/Rekrutierung	14 Nähere Angaben über den Zeitraum der Studienaufnahme der Probanden/Patienten und der Nachbeobachtung.	
Patientencharakteristika zu Studienbeginn (baseline data)	15 Demografische und klinische Charakteristika aller Gruppen.	
Anzahl der ausgewerteten Probanden/Patienten	16 Anzahl der Probanden/Patienten (Nenner) in jeder Gruppe, die in die entsprechende Analyse eingeschlossen wurden und Angabe, ob es sich dabei um eine „Intention-to-Treat- Analyse handelt. Wenn möglich, Angabe der Ergebnisse in absoluten Zahlen (z.B. 10 von 20, nicht 50%).	
Ergebnisse und Schätzmethode	17 Zusammenfassung der Ergebnisse aller primären und sekundären Zielkriterien für jede Gruppe und die geschätzte Effektgröße sowie ihre Präzision (z.B. 95%-Konfidenzintervall).	
Zusätzliche Analysen	18 Angabe von weiteren Tests, insbesondere von Subgruppenanalysen und adjustierte Analysen (mit Erklärung, ob sie vorher geplant waren oder nachträglich durchgeführt wurden).	
Unerwünschte Wirkungen	19 Angabe aller wichtigen unerwünschten Wirkungen oder Nebenwirkungen innerhalb jeder Behandlungsgruppe.	
Diskussion		
Interpretation	20 Interpretation der Ergebnisse unter Berücksichtigung der Studienhypothesen, möglicher Ursachen von Verzerrungen („Bias“) sowie Problemen durch multiples Testen und multiple Zielkriterien.	
Generalisierbarkeit	21 Generalisierbarkeit der Studienergebnisse (externe Validität).	
Bewertung der Evidenz	22 Allgemeine Interpretation der Ergebnisse unter Berücksichtigung des aktuellen Forschungsstandes und anderer Publikationen zur untersuchten Fragestellung.	

Abbildung 3: CONSORT-Checkliste zur Publikation randomisierter Studien (<http://www.cochrane.de/deutsch/cccons1.htm> [166])

Genau an dem Punkt der Transparenz setzt das EBM-Schema nach Sackett ein, das für die Evaluation einer Studie erstellt wurde. Eine RCT kann nur dann komplett beurteilt werden, wenn die zu einer Bewertung wichtigen Informationen darin enthalten sind. Die Checkliste des CONSORT-Statements ist somit eine Hilfe für Autoren, denn bisher musste derjenige, der über die Qualität einer Studie urteilen wollte, bei fehlenden Angaben über das jeweilige Qualitätsmerkmal einer Studie dieses als nicht ordnungsgemäß durchgeführt beurteilen. Eine Studie, in der alle CONSORT-Forderungen erfüllt sind, wird dem Problem, dass der Leser spekulieren muss, jedoch aus dem Weg gehen (Altman 2001, S.686 [48]), und Moher et al. 2001 [109] konnten mittlerweile beweisen, dass das CONSORT-Statement tatsächlich zur Qualitätsverbesserung von Studienberichten beiträgt (Moher 2001, S.1992f. [109]).

Die zweite neben der Checkliste angebotene Hilfe des CONSORT-Statements ist ein Flussdiagramm, das exemplarisch aufzeigt, wie der Verlauf von Patienten durch eine Studie dargestellt werden sollte, um dem Leser alle für eine vollständige Beurteilung notwendigen Informationen zu vermitteln (Abbildung 4) (Falck-Ytter 2002 [165]). Auch dieses ist elementarer Bestandteil eines guten Berichts, denn falls der Patientenfluss- und -verlust während einer Studie nicht transparent ist, ist eine klare Aussage über die Ursachen für am Ende fehlende Patienten und damit eine objektive Bewertung der Validität nicht möglich.

In der vorliegenden Arbeit wurde die Checkliste des CONSORT-Statements für die Anfertigung eines Fragebogens verwendet, der speziell für die Datenaufnahme aus MR-gestützten MS-Therapiestudien aufgestellt wurde (siehe Abbildung 2). Dieser Fragebogen erfasst zunächst allgemeine und die für MS wichtigen Angaben und anschließend wie in der CONSORT-Checkliste die einzelnen Aspekte nach der in RCTs üblichen Gliederung in Abstract, Einleitung, Methoden, Resultate und Diskussion. Nach Abschluss der Durchsicht aller eingeschlossenen Studien erfolgte die Auswertung der Fragebögen im Hinblick darauf, welche Aspekte in modernen MS-Therapiestudien meist beachtet werden und auf welchen Gebieten Verbesserungen nötig sind. Außerdem waren Teile der Bögen Grundlage für die Anfertigung des Reviews über Therapiemethoden bei MS.

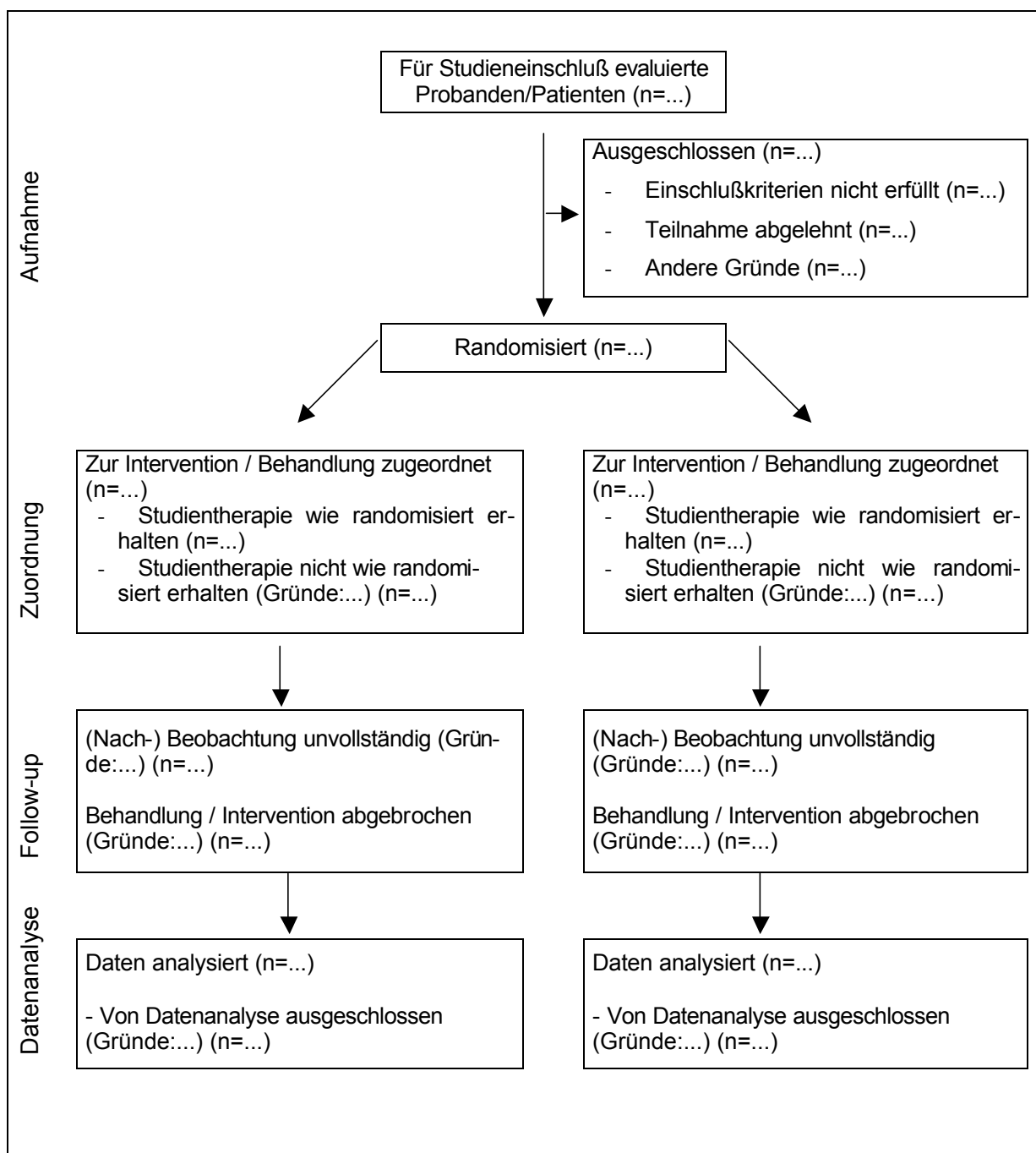


Abbildung 4:: Beispiel eines Flussdiagramms für die ein- und ausgeschlossenen Probanden/Patienten im Verlauf einer randomisierten Studie (Aufnahme, Zuordnung, Nachbeobachtung und Datenanalyse). (<http://www.cochrane.de/deutsch/cccons1.htm> [164])

2.4. EBM-Bewertungsschema nach Sackett

Für die Untersuchung und Bewertung der eingeschlossenen Studien wurde das EBM-Bewertungsschema nach Sackett und Kollegen angewandt (Sackett 1999, Ss. 72-76, 105-111, 129-138 [134]). Dieses Schema besteht aus drei Teilen. Der erste Teil befasst sich mit der Bewertung der Validität, der zweite mit der Bewertung der Wichtigkeit und der dritte mit der Überprüfung der Anwendbarkeit auf den Patienten, wobei die Schritte eins und zwei vertauscht werden dürfen. Es macht keinen Unterschied, ob zunächst die Validität oder die Wichtigkeit beurteilt werden: Sollte die Validität einer Studie nicht vorhanden sein, muss die Beurteilung der Wichtigkeit nicht mehr durchgeführt werden, und falls das Ergebnis einer Studie nicht wichtig ist, ist es irrelevant, ob die Studie valide ist.

Abbildung 5 zeigt das Schema in seiner ursprünglichen Form. Alle in die Auswertung eingeschlossenen Studien wurden zunächst damit bewertet. Für einen zweiten Durchgang wurde dieser Bewertungsplan, wie in Abschnitt 2.5 dargestellt, modifiziert, um eine möglichst objektive Beurteilung der eingeschlossenen Studien zu gewährleisten. Die Resultate der Bewertung finden sich im Ergebnisteil dieser Arbeit. Zunächst werden jedoch in den Abschnitten 2.4.1 bis 2.4.3 die Grundzüge des EBM-Schemas nach Sackett erläutert.

Bewertung einer Therapiestudie	
Untersuchte Studie:	
Zeitschrift	
Jahr, Nummer, Seiten	
Artikel	
Betrifft Arbeitsthese	
I. Ist die Evidence zu der gefundenen Behandlung valide?	
1. Wurden die Patienten den Behandlungen randomisiert zugeordnet und wurde die Randomisierungsliste geheimgehalten ?	
2. Wurden bei der Gesamtbeurteilung der Studie alle darin aufgenommenen Patienten berücksichtigt ?	
3. Wurde eine „ Intention-to-treat “-Analyse durchgeführt?	

Abbildung 5: Bewertungsschema nach D. Sackett (Sackett 1999 [134]) (fortgesetzt auf der folgenden Seite)

4. Waren Patienten und Ärzte gegenüber der durchgeführten Behandlung verblindet ?	
5. Wurden die Gruppen, abgesehen von der experimentellen Behandlung, gleich behandelt ?	
6. Waren die Gruppen zu Beginn der Studie einander ähnlich ?	

II. Ist die Evidence zu einer Behandlung wichtig?

<p>1a. Finden der klinisch nützlichsten Aussage einer validen Therapiestudie</p> <p>1b. Vergleich der Ergebnisse mit den Ergebnissen anderer Behandlungen von anderen Erkrankungen</p>	
<p>2a. Berechnung der Relativen Risikoreduktion (RRR):</p> <p>$RRR = (CER - EER)/CER$ (CER=control event rate = Ereignisrate der Vergleichspatienten, EER = experimental event rate = Ereignisrate der Verumpatienten)</p> <p>2b. Berechnung der Absoluten Risikoreduktion (ARR):</p> <p>$ARR = CER - EER$</p>	
<p>3a. Berechnung der Anzahl an Patienten, die behandelt werden müssen, um ein negatives Ergebnis zu verhindern (Number needed to be treated= NNT):</p> <p>$NNT = 1/ARR$</p> <p>3b. Berechnung der NNT aus den Chancenverhältnissen (Odds Ratios= OR), falls NNT nicht direkt angegeben und stattdessen OR:</p> <p>$NNT = \frac{1 - (PEER * (1 - OR))}{(1 - PEER) * PEER * (1 - OR)}$</p> <p>(PEER = patient expected event rate = beim Patienten zu erwartende Ereignisrate)</p>	

Abbildung 5: Fortsetzung Bewertungsschema nach D. Sackett (Sackett 1999 [134])
(fortgesetzt auf der folgenden Seite)

III. <u>Ist die valide und wichtige Evidence zur Therapie auf die medizinische Versorgung des Patienten anwendbar?</u>	
1. Gelten die Ergebnisse der Studie auch für unseren Patienten ?	
2. Welche Auswirkungen hat die Behandlung auf spezielle Untergruppen ? Könnte es sein, dass mein Patient einer Untergruppe angehört, die u.U. nicht von einer Behandlung profitiert oder der diese Behandlung sogar schadet?	
<p>3. Wie groß wäre der mögliche Nutzen der Therapie für meinen Patienten?</p> <p><u>Für die Beantwortung dieser Frage gibt es zwei Möglichkeiten:</u></p> <p>3.1. 3.1.1. Anfälligkeit des Patienten a) Wenn die Merkmale des Patienten mit den Merkmalen der Studienpatienten übereinstimmen: Übernahme des entsprechenden Wertes für die Anfälligkeit des Patienten b) Falls die Werte nicht übereinstimmen: Literaturrecherche, um herauszufinden, welche Prognose der Patient ohne Behandlung hätte.</p> <p>3.1.2. Multiplikation des erhaltenen Prozentwertes mit der RRR => Resultat ARR => Umrechnung in NNT (alternativ Anwendung des Nomogramms im Anhang, Abbildung 4)</p> <p>3.2. NNT aus der Studie geteilt durch den angenommenen Wert (entspr. dem auf meinem Patienten umgerechneten Wert der Studien-Placebogruppe für die Anfälligkeit des Patienten): NNT / F</p>	
4. Vergleich der Wertvorstellungen und Präferenzen meines Patienten mit dem Therapieplan und dessen Folgen:	

Abbildung 5: Fortsetzung Bewertungsschema nach D. Sackett (Sackett 1999 [134])

2.4.1. Validität

Der erste Abschnitt der Bewertung nach Sackett befasst sich mit der Validität einer Studie. Diese wird in sechs Schritten untersucht: Randomisierung, Einschluss aller Patienten, Intention-to-treat, Verblindung, Gleichbehandlung und Ähnlichkeit der Behandlungsgruppen.

2.4.1.1. Randomisierung

Der erste Schritt ist die Frage nach der Randomisierung. Oftmals ist diese Frage schon in der Überschrift oder im Abstract beantwortet, denn ob randomisiert wurde, ist eines der Hauptkriterien für die Validität einer Studie. Die Zufallsverteilung von Patienten ist wichtig, weil nur so die unbekannt Risiken gleichmäßig auf die verschiedenen Behandlungsgruppen verteilt werden können. Dabei liegt die Betonung auf „unbekannt“, denn bei bekannten Risiken besteht die Möglichkeit, die Patienten den jeweiligen Gruppen zuzuordnen. Es könnte zum Beispiel sein, dass Männer ein höheres Risiko für eine in einer Studie untersuchte bestimmte Krankheit tragen. Dann müssten in einer zweiarmigen Studie, in der beispielsweise 20 Männer und 20 Frauen aufgenommen werden, deren sonstige Risiken aber völlig ausgewogen sind, je zehn Männer und zehn Frauen in jede Behandlungsgruppe eingeschlossen werden. Nur ist dies in der Realität so nicht durchführbar, denn fast immer existieren unbekannte Risiken, die es verhindern, die Patienten ihren Merkmalen entsprechend gleich zu verteilen, ohne dabei einen Fehler hinsichtlich der Verteilung der Risiken zu machen. Deshalb gilt die randomisierte kontrollierte Studie berechtigterweise als „Goldstandard“ (Schulz 1997, S.569 [142]) für die Durchführung klinischer Untersuchungen, und nicht-randomisierte Studien sollten nur dann für die Entscheidungsfindung hinsichtlich einer Therapie für einen bestimmten Patienten hergenommen werden, wenn keine andere Evidence zur Verfügung steht (Guyatt 1993, S.2599 [86]). Die Frage nach der Randomisation ist im Schema nach Sackett (Sackett 1999, Ss. 72-76, 105-111, 129-138 [134]) nochmals unterteilt in zwei Aspekte.

Der erste Teil prüft, ob eine Randomisierung überhaupt stattfand. Dabei muss die Zuteilung der Patienten zu den Behandlungsgruppen mit einer Methode vorgenommen werden, „(...) die dem Werfen einer Münze entspricht“ (Sackett 1999, S.73 [134]). In der Praxis wird zunächst von einer an der Studie unbeteiligten Person, zum Beispiel mithilfe eines Computerprogramms, durch Würfeln oder durch Ziehen von Karten mit Nummern, eine Liste erstellt, auf der die Zuteilung der einzelnen Patienten zu den verschiedenen Behandlungsgruppen genau aufgeführt ist. Mit diesem Verfahren werden die Vergleichsgruppen „(...) hinsichtlich der prognostischen Faktoren angeglichen, (...) denn wenn diese nicht einheitlich in den Behandlungsgruppen verteilt wären, könnten sie die Wirkungen der Therapie übertreiben, aufheben oder ihr sogar entgegenwirken.“ (Sackett 1999, S.73f. [134]) Falls die ungleich verteilten Faktoren zu einem falsch positiven Ergebnis der Behandlung führten, würde dadurch die Wirkung der Therapie übertrieben und damit die falsche Schlussfolgerung gezogen, die untersuchte Behandlung sei nützlich. Ebenso würde bei einer Beeinflussung in Richtung eines falsch negativen Ergebnisses die Therapiewirkung verschleiert und eine möglicherweise vielversprechende Behandlungsform nicht als wirksam erkannt (Sackett 1999, S.73f. [134]).

Im zweiten Teil des Abschnitts „Randomisierung“ wird nach der Geheimhaltung gefragt. Es ist wichtig, dass die Aufnahme der Patienten in die Studie und damit die Verteilung auf die Therapiegruppen ohne die Information über die zugewiesene Behandlung geschieht (Schulz 1997, S.570 [142]). Dies ist deshalb von besonderer Bedeutung, weil damit keine der beteiligten Personen das Gleichgewicht zwischen den Gruppen wissentlich oder unwissentlich stören kann (Sackett 1999, S.74 [134]).

Außerdem führt es zur Überschätzung des Behandlungseffekts, wenn die Methode der Geheimhaltung inadäquat oder unklar ist (Huston 1996, S. 1277 [91]). Deshalb sollen hier exemplarisch die zwei gängigsten Methoden in ihrem Ablauf kurz beschrieben werden. Die erste Möglichkeit ist die Arbeit mit versiegelten Umschlägen. Nach Erstellung des Zuteilungsschemas werden dabei die Ergebnisse auf nummerierte Umschläge verteilt, die anschließend zugeklebt werden. Es enthält dann jeder Umschlag genau eine Zuteilung, wobei das entsprechende Kuvert erst geöffnet werden darf, wenn dokumentiert ist, dass der zu behandelnde Patient sicher an der Studie teilnimmt. Diese Regel soll eine Geheimhaltung der Randomisierung gewährleisten. Die Kuvert-Methode hat jedoch unter Umständen den Nachteil einer Manipulation. So können Umschläge einfach geöffnet oder dünne Kuverts gegen das Licht gehalten werden (Schulz 1997, S.570 [142]). Falls keine Überwachung stattfindet, kann also ein Arzt oder ein anderer Beteiligter, der mit der Zuteilung eines bestimmten Patienten nicht einverstanden ist, einen anderen Umschlag wählen, der dann vielleicht eine seiner Ansicht nach bessere Einordnung des Patienten in eine Behandlungsgruppe bietet. Tendenziell wird die experimentelle Behandlung in diesem Fall eher den prognostisch etwas besseren Patienten zugeteilt, was zur Ergebnisverfälschung führt. Eine Therapie wird dann womöglich als wirksam angesehen, obwohl dies in Wahrheit nicht zutrifft.

Ein weiteres, aufwändigeres, aber sichereres Vorgehen im Hinblick auf die Geheimhaltung ist eine zentral gesteuerte Randomisation. Bei dieser Form verbleibt die erstellte Liste bei einer von der Studie vollkommen unabhängigen Institution. Der Studienleiter muss dann bei Einschluss eines Patienten in die entsprechende Studie unter Nennung des Patientennamens mit dieser Institution Rücksprache halten und erhält anschließend die Gruppenzugehörigkeit des entsprechenden Patienten zugewiesen. Eine Manipulation ist mit dieser Methode praktisch ausgeschlossen oder zumindest höchst unwahrscheinlich. Demnach ist es als ein Qualitätskriterium zu werten, wenn in einer Veröffentlichung von einer zentralen Randomisation gesprochen wird (DerSimonian 1982, S.1334 [69]). Aus Gründen der Transparenz einer Publikation ist es jedoch noch besser, wenn das Vorgehen bei der Zuordnung der Patienten zu den Behandlungsgruppen genau erklärt wird (Moher 2001 [110], S.659; Begg 1996 [64], S.638; Moher 2001 [107], S.1989; Falck-Ytter 2002 [165]): Es sollte also beschrieben werden, welche Methode verwendet, wie diese genau durchgeführt und geheimgehalten wurde.

Für die vorliegende Arbeit wurde bei der Bewertung der Randomisation in den untersuchten Studien wie folgt vorgegangen. Zunächst wurde entsprechend dem CONSORT-Statement untersucht, ob die Methode der Randomisation exakt beschrieben war, das heißt, ob in der jeweils untersuchten Studie geschildert war, wie die Randomisationssequenz hergestellt, wie sie geheimgehalten und wie die Zufallszuteilung schließlich durchgeführt wurde. Außerdem wurde der Zeitpunkt der Randomisation bewertet und überprüft, ob die Patienten nicht zu früh oder zu spät den Randomisationsvorgang durchliefen. Anschließend erfolgte die Beantwortung der zweigeteilten Frage des Sackett-Schemas. Schließlich wurden die Ergebnisse anhand des neu entwickelten objektivierte Bewertungssystems nach Hassler, auf das im Abschnitt 2.5 dieser Arbeit genauer eingegangen wird, evaluiert und damit die Bewertung abgeschlossen.

2.4.1.2. Berücksichtigung aller eingeschlossenen Patienten in der Gesamtwertung

Im Anschluss an die Randomisation folgt als zweiter Schritt für die Überprüfung der Validität mithilfe des Sackett-Schemas die Frage, ob alle zu Beginn in die Studie aufgenommenen Patienten am Ende in die Gesamtwertung eingeschlossen wurden. Dafür vergleicht man die Zahl der Patienten zu Beginn mit der zum Schluss, wobei diese idealer Weise gleich sein sollte. Falls nämlich Patienten mit negativem Endergebnis vorzeitig die Studie beenden, kann das Ergebnis der Studie überschätzt werden. In der Regel ist ein Verlust von bis zu 10% noch akzeptabel, während eine Studie bei einem Verlust von mehr als 20% nicht mehr als voll valide gelten kann. Um einer Überbewertung der Studienergebnisse vorzubeugen, „(...) sollte es den Autoren möglich sein, bei sämtlichen verlorengegangenen Patienten den schlimmstmöglichen Fall ('Worst-Case'-Szenario) anzunehmen (...)“ (Sackett 1999, S.75 [134]). Das bedeutet, dass alle vorzeitig ausgeschiedenen Patienten mit positivem Testresultat in der abschließenden Bewertung auch als mit negativem Resultat ausgeschieden betrachtet werden können ohne dabei das Gesamtergebnis zu verfälschen. Umgekehrt sollte bei Patienten mit negativem Resultat zum Zeitpunkt des Ausscheidens aus der Studie auch ein positives Testergebnis ohne Veränderung der Gesamtaussage der Studie angenommen werden können (Sackett 1999, S.75 [134]).

Für die vorliegende Arbeit dienten die in der jeweils bewerteten Studie veröffentlichten Patientenzahlen als Grundlage für eine Berechnung. Auf im Anhang erwähnte Sekundärliteratur wurde zurückgegriffen, wenn die Daten der Primärliteratur zu knapp oder nicht eindeutig waren. Bei Diskrepanzen zwischen Primär- und Sekundärliteratur wurde der niedrigere von beiden Prozentwerten für die abschließende Beurteilung mit dem objektivierten Bewertungssys-

tem (siehe Abschnitt 2.5) verwendet. Tabelle 3 im Anhang zeigt eine Auflistung der für die Bewertung herangezogenen Primär- oder Schwesterstudien.

2.4.1.3. Intention-to-treat-Analyse

Im Zusammenhang mit der Auswertung der Patienten ist es nicht nur wichtig, dass möglichst wenige im Studienverlauf ausgeschieden sind, sondern auch, dass die Patienten in den Gruppen ausgewertet wurden, denen sie zugeordnet waren. Die sogenannte Intention-to-treat-Analyse (ITT) hat dabei im Wesentlichen zwei Ziele: 1.) Die Randomisation, die bekannte und unbekannte Risiken gleich verteilt (siehe Abschnitt 2.4.1.1.), soll aufrechterhalten werden; 2.) ITT berücksichtigt Non-Compliance und Abweichungen vom Studienprotokoll (Hollis 1999, S.670f. [90]). Falls ein Patient zum Beispiel in die Verum-Gruppe randomisiert worden war, aber aufgrund eines Fehlers das Placebo oder auch mangels Compliance gar keine Medikamente einnahm, muss er trotzdem in der Verum-Gruppe ausgewertet werden. Denn „es konnte (...) wiederholt nachgewiesen werden, dass Patienten, die ihre Medizin einnehmen bzw. nicht einnehmen, zu sehr unterschiedlichen Ergebnissen kommen, auch wenn es sich bei der Studienmedikation, die ihnen verordnet wurde, um ein Placebo handelt“ (Sackett 1999, S.76 [134]). Außerdem können bei heutigen modernen Therapien, insbesondere bei Krankheiten, für die es schon eine etablierte Standardtherapie gibt, Veränderungen, hervorgerufen durch biologische Unterschiede, größer sein als die durch die therapeutische Intervention hervorgerufenen (Porzsolt 2000 [124], Porzsolt 2001 [125]). Jeder in eine Studie eingeschlossene Patient muss also in Form einer „Intention-to-treat“-Analyse (ITT), unabhängig von seiner tatsächlichen Behandlung, genau in der Gruppe ausgewertet werden, in die er zu Beginn randomisiert worden war. Es wäre ein großer Fehler, Patienten mit Fehlmedikation komplett aus den Ergebnissen heraus zu nehmen. Es kann aber, abhängig von der jeweiligen Fragestellung, notwendig sein, parallel zur ITT eine sogenannte „Per-Protocol“-Auswertung durchzuführen, die unter Umständen zusätzliche Erkenntnisse liefern kann (Schäfer 1999, S.153 [138]).

Für die Studienbewertung in der vorliegenden Arbeit wurde nach Anzeichen gesucht, ob eine ITT durchgeführt wurde oder nicht. Die Erwähnung einer ITT alleine kann zwar schon einen Hinweis darauf geben, es müssen jedoch auch alle weiteren Kriterien für eine positive Bewertung erfüllt sein. Falls nämlich an anderer Stelle im Text deutlich wird, dass Patienten, die nicht die vorgesehene Studienmedikation erhielten, aus der Bewertung herausgelassen oder doch „per protocol“ ausgewertet wurden, kann die Frage nach einer ITT nur negativ beantwortet werden. Ähnlich der Beurteilung der Randomisation wurden zunächst die nach CONSORT relevanten Punkte überprüft, anschließend mithilfe des Sackett-Schemas die Frage

Frage nach einer ITT beantwortet, und schließlich erfolgte die Punktvergabe im objektivierten Bewertungssystem nach Hassler.

2.4.1.4. Verblindung

Als nächster für die Validität einer Studie wichtige Punkt ist die Verblindung zu nennen. Wenn möglich, sollte eine doppelblinde oder gar dreifachblinde Studienform verwendet werden. Doppelblind bedeutet, dass sowohl Patient als auch behandelnder Arzt über die Studienmedikation nicht informiert sind (Schäfer 1999, S.153 [138]). Eine besondere Form der Doppelverblindung ist die zusätzliche Unkenntnis des die Diagnostik durchführenden Arztes (Schäfer 1999, S.153 [138]). Bei sogenannten dreifachblinden Studien ist außerdem der Biometriker verblindet, sodass dieser keine Möglichkeit hat, aus den Studienergebnissen Rückschlüsse auf die erhaltene Medikation zu ziehen (Schäfer 1999, S.153 [138]). Die dreifachverblindete Studienform kann noch besser als die doppelblinde dazu beitragen, einen Bias (=systematischer Fehler (Perleth 1999, S.101 [121])), zu verhindern. Doch auch die Doppelverblindung ist ein sehr gutes Mittel, um durch die Kenntnis der zugewiesenen Studienmedikation entstehende Fehler im Outcome, zum Beispiel aufgrund des Placeboeffektes, auszuschließen.

Falls jedoch dreifach- oder doppelblinde Studien nicht möglich sind, sollte zumindest eine Einfachverblindung erfolgen. Das bedeutet, dass nur der Patient über seine Therapie nicht informiert ist. Außerdem sollte in einem derartigen Fall versucht werden, eine unabhängige Bewertung der Ergebnisse des Patienten zu erhalten, also einen mit der Therapie nicht befassten Arzt mit der Diagnostik zu beauftragen (Schäfer 1999, S.153 [138]). Zum Beispiel gäbe es die Möglichkeit, in Studien, mit denen die Wirkung von Protonenpumpenhemmern auf die Magenerkrankung untersucht wird, einen vorher nicht involvierten Arzt einzusetzen, der bei den Studienpatienten eine Gastroskopie durchführt und damit den Erfolg der Therapie beurteilt. Eine Einfachverblindung sollte aber nur eingesetzt werden, wenn eine Mehrfachverblindung nicht möglich ist und kann allenfalls eine Ersatzlösung sein.

Es gibt jedoch auch Studien, in denen eine Verblindung nicht durchführbar ist. Unverblindete Studien können mit Einschränkungen als Evidence für eine Therapiemethode angesehen werden, wenn es überhaupt keine Möglichkeit gibt, eine Verblindung vorzunehmen. Dies ist jedoch selten der Fall, denn selbst bei vielen chirurgischen Eingriffen kann man als Placeboanwendung zum Beispiel einen Hautschnitt setzen und damit die Therapie zumindest für den Patienten verschleiern. Falls eine Verblindung eigentlich denkbar wäre, aber nicht durchgeführt wurde und neben der Studie noch keinen weiteren Untersuchungen zu dem entsprechenden Thema existieren, kann eine derartige Veröffentlichung immerhin als vorläufige Evi-

dence bis zum Vorhandensein verblindeter Studien angesehen werden. Als Grundlage für Behandlungen sollte eine solche Studie aber nicht dienen.

In der vorliegenden Arbeit wurde eine Verblindung von Patienten, diagnostizierenden und therapeutisch tätigen Ärzten geprüft. Falls darüber keine direkten Angaben vorlagen, wurde versucht, aus weiteren Beschreibungen im Text die notwendigen Informationen zu erhalten. War auch das nicht erfolgreich, wurde auf Angaben in Erst- oder Zweitveröffentlichungen zu den bewerteten Studien zurückgegriffen. Wenn keine Informationen über eine Verblindung vorlagen, wurde die entsprechende Studie als „nicht verblindet“ bewertet. Die gewonnenen Informationen wurden schließlich in die Punktvorgabe des objektivierte Bewertungssystems übersetzt.

2.4.1.5. Gleichbehandlung

Ein nicht so deutlich hervortretendes und oft als selbstverständlich angesehenes Kriterium der Validität ist die Gleichbehandlung der verschiedenen Behandlungsgruppen. Es ist wichtig, dass beide Gruppen neben der experimentellen Therapie ansonsten völlig identisch behandelt werden. Eine weitere Intervention oder ein bestimmtes diagnostisches Verfahren darf nicht allein einer der zwei oder mehreren zu vergleichenden Therapiegruppen vorbehalten sein, weil es sonst zu Verfälschungen des Ergebnisses kommen kann, sei es durch einen tatsächlichen Effekt einer zusätzlichen Therapie oder durch einen Placeboeffekt. In beiden Fällen wäre nicht mehr eindeutig feststellbar, durch welche der angewandten Methoden Unterschiede im Outcome erklärbar wären.

Ein derartiges Problem träte zum Beispiel auf, wenn in einer Studie von Patienten mit koronarer Herzkrankheit (KHK) die Vergleichsgruppe zunächst vor perkutaner koronarer Intervention eine Vorbehandlung mit Aspirin erhielte, während die Verum-Patienten, als Studienmedikation eine Clopidogrelnachbehandlung erhaltend, nicht vorbehandelt werden. Man könnte in einem derartigen Fall nicht entscheiden, wodurch ein möglicherweise beobachteter Effekt zustande kommt: Ist es generell die Vor- oder Nachbehandlung mit einem gerinnungshemmenden Medikament? Wie groß ist ein möglicher Erfolg der Studienmedikation im Vergleich zur Standardbehandlung? Der Effekt wäre nicht eindeutig der Aspirinvorbehandlung oder der Clopidogrelnachbehandlung zuzuschreiben und die ~~Studienresultate~~ ^{Studienresultate} zum Beispiel wurde in der vorliegenden Arbeit untersucht, ob die Ergebnisse eindeutig der Studientherapie zuzuordnen waren oder ob andere Faktoren das Outcome der Patienten beeinflusst haben könnten. Die Gleichbehandlung wurde geprüft und mithilfe des objektivierte Bewertungssystems nach Hassler beurteilt.

2.4.1.6. Ähnlichkeit der Patienten

Nach der Überprüfung der Gleichbehandlung ist die Untersuchung der Ähnlichkeit der Vergleichsgruppen in einer Studie das letzte Kriterium, das für die Evaluation der Validität von Bedeutung ist (Sackett 1999, S.76 [134]). In der Regel kann die Übereinstimmung der Merkmale am besten beurteilt werden, wenn in einer Tabelle, ähnlich wie in Jacobs 2000, S.901 [13], die für eine Krankheit und deren Therapie relevanten Eigenschaften der eingeschlossenen Patienten gegenübergestellt sind (siehe Abbildung 6 im Anhang). Vergleichskriterien sind zum Beispiel die Anzahl der Patienten pro Behandlungsgruppe, das durchschnittliche Alter der Patienten, das Verhältnis von Frauen zu Männern, die Schwere der untersuchten Krankheit, Übereinstimmung in der vorliegenden Krankheitsentität, bestehende Begleit- oder Vorerkrankungen und einiges mehr. Wie wichtig das einzelne Kriterium für die Ähnlichkeit der Patienten ist, muss in jeder Studie, abhängig von der untersuchten Krankheit, neu definiert werden. Selbstverständlich sollte die Übereinstimmung zusätzlich zu den aufgezählten Kriterien auch in anderen wichtigen Gegebenheiten untersucht werden. Bei peripherer arterieller Verschlusskrankheit (pAVK) ist zum Beispiel das Gewicht des Patienten zusammen mit seiner Größe relevant, weil mit diesen beiden Werten der bei pAVK prognostisch wichtige Body Mass Index (BMI) errechnet werden kann. Falls nun in einer von zwei Vergleichsgruppen einer pAVK-Therapiestudie alle Merkmale bis auf den BMI übereinstimmen, dieser aber signifikant unterschiedlich ist, sind die Ergebnisse der Studie nur noch eingeschränkt zu verwerten. In einer anderen Studie hingegen, zum Beispiel zur MS, ist der BMI eine eher unwichtige Kenngröße. Es ist also notwendig, Wert auf eine vollständige Dar- und Gegenüberstellung aller relevanten Parameter eines Patientenkollektivs zu legen, um die Ergebnisse einer Studie hinsichtlich ihrer Relevanz einschätzen zu können.

Für die Beurteilung der Ähnlichkeit zweier Vergleichsgruppen wurde in der vorliegenden Arbeit zunächst überprüft, ob in den untersuchten Studien eine Tabelle mit Gegenüberstellung aller relevanten Daten vorhanden war und anschließend, ob die erstellten Gruppen in der Ähnlichkeit übereinstimmten. Mangelhafte Transparenz, das heißt Fehlen einer Tabelle der Patientendaten bei gleichzeitig mangelnder Beschreibung der Patienten im Text, führte im objektivierten Bewertungssystem nach Hassler ebenso zu Punktabzug wie eine Ungleichheit der Vergleichsgruppen in den jeweils für die entsprechende Studie wichtigen Kriterien.

2.4.2. Wichtigkeit

Neben der Validität einer Studie spielt, wie oben erwähnt, die Wichtigkeit ebenfalls eine große Rolle, wobei unerheblich ist, ob bei der Überprüfung mit der Validität oder der Wichtigkeit

begonnen wird (Sackett 1999, S.72 [134]). Der zweite Teil des EBM-Bewertungsbogens nach Sackett überprüft den Nutzen und das Ausmaß der Wirkungen einer in der bewerteten Studie durchgeführten Behandlung (Sackett 1999, S.105 [134]). Zunächst muss man dazu die klinisch nützlichste Aussage einer Therapiestudie finden und anschließend die „(...) Ergebnisse mit den Ergebnissen anderer Behandlungen von anderen Erkrankungen vergleichen“ (Sackett 1999, S.105 [134]). Dazu existieren verschiedene miteinander zusammenhängende Kennzahlen, mit deren Hilfe man die Wirkung abschätzen kann. Die erste Kennzahl, die hier beschrieben werden soll, ist die Risikoreduktion.

Diese begründet sich darauf, dass jeder an einer Studie teilnehmende Patient ein bestimmtes Risiko des Auftretens eines Ereignisses hat. Sollte die untersuchte Therapie einen Nutzen dahingehend haben, dass bei den Patienten der experimentellen Gruppe ein unerwünschtes Ereignis seltener vorkommt als in der Kontrollgruppe, spricht man von Risikoreduktion. Man unterscheidet die absolute Risikoreduktion (ARR) und die relative Risikoreduktion (RRR). In der RRR wird der Risikounterschied der beiden Behandlungsgruppen ins Verhältnis zu dem Risiko der Kontrollgruppe gesetzt. Die entsprechende Formel stellt sich folgendermaßen dar: $RRR = (CER - EER)/CER$ (CER = control event rate, also die Ereignisrate in der Vergleichsgruppe, EER = experimental event rate, also die Ereignisrate in der Experimentalgruppe). Die RRR macht jedoch keinen Unterschied zwischen großen und kleinen Therapiewirkungen. Deshalb ist es wichtig, eine zweite Kenngröße zu haben, in die das Ausgangsrisiko der Patienten mit einfließt und an der man die tatsächliche Risikoverminderung abschätzen kann, die ARR. Diese lässt sich wie folgt berechnen: $ARR = CER - EER$. Am Ergebnis wird der tatsächliche Risikounterschied der verglichenen Gruppen deutlich. Angenommen, die experimentelle Gruppe hätte eine Ereignisrate für das Auftreten eines unerwünschten Ereignisses von 10% und die Kontrollgruppe von 50%. Die ARR wäre also $50\% - 10\% = 40\%$, während die RRR als Ergebnis $(50\% - 10\%)/50\% = 80\%$ hätte. Sollte der Behandlungseffekt jedoch nur jeweils 1/100 sein, kämen wir auf folgende Ergebnisse für ARR und RRR: $ARR = 0,5\% - 0,1\% = 0,4\%$; $RRR = (0,5\% - 0,1\%)/0,5\% = 80\%$. Die Minimierung des Therapieeffekts auf 1/100 wird deutlich in der ARR, stellt sich aber in der RRR überhaupt nicht dar!

Natürlich lässt sich, falls mit einer Therapie nicht ein negatives Ereignis verhindert sondern herbeigeführt wird, auch eine absolute Risikozunahme, die ARI (=absolute risk increase) und eine relative Risikozunahme, die RRI (=relative risk increase), berechnen. Die Vorgehensweise erfolgt analog der Berechnung der ARR/RRR.

Für die klinische Anwendung sind sowohl die ARR/ARI als auch die RRR/RRI alleine nicht besonders aussagekräftig. Deshalb hat man eine weitere Kenngröße eingeführt. Die Zahl der Patienten, die behandelt werden müssen, um bei einem das Auftreten des unerwünschten Ereignisses zu verhindern, drückt sich aus in der „number needed to treat“, der NNT. Diese

berechnet sich aus dem Kehrwert der ARR mit folgender Formel: $NNT = 1/ARR$. Entsprechend der ARI und der RRI bei ungünstigem Therapieergebnis existiert auch analog zur NNT eine Bezeichnung, die „number needed to harm“ oder NNH. Falls die Therapie also für den Patienten im Vergleich zur Kontrollgruppe schädlich ist, drückt man damit die Anzahl der Patienten aus, die behandelt werden müssten, um bei einem zusätzlich das unerwünschte Ereignis hervor zu rufen.

Nach Berechnung der NNT/NNH folgt der zweite Schritt in der Bewertung der Wichtigkeit einer Therapiestudie. Dies ist der Vergleich der errechneten NNT mit den NNTs anderer Therapien bei anderen Erkrankungen. Abbildung 7 im Anhang zeigt verschiedene NNTs diverser Krankheiten auf, um eine Gegenüberstellung zu erleichtern. So lässt sich einschätzen, wie groß der in einer Studie gefundene Behandlungseffekt wirklich ist.

Eine Größe, die zwar nicht in die Berechnung der NNT mit einfließt, aber dennoch in die Betrachtung des Behandlungserfolgs mit eingebunden werden muss, ist die Dauer der Therapie, denn falls ein Effekt zum Beispiel schon nach einem Monat der Behandlung auftritt, ist das günstiger als ein Therapieerfolg erst nach fünf Jahren.

Voraussetzung für die Berechnung der Kennzahlen RRR, ARR und NNT in der vorliegenden Arbeit waren folgende zwei Kriterien.

1. Die Studie musste auf dem Bewertungsschema nach Sackett mindestens eingeschränkt valide sein bzw. mit dem neu erstellten objektiven Evaluierungsbogen nach Hassler (siehe Gliederungspunkt 2.5), der die Grade 0 (=nicht valide) bis 4 (=voll valide) umfasst, Grad 2 der Validität erreicht haben. Es macht keinen Sinn, in einer Studie, auf deren Validität man nicht vertrauen kann, die Wichtigkeit der Daten zu überprüfen.
2. In den untersuchten Studien mussten die Werte entsprechend transparent und nachvollziehbar sein, sodass Outcomes eindeutig bestimmten Patienten zuzuordnen waren.

Als bedeutsame Kriterien der einzelnen Studien wurden in dieser Arbeit zum Einen Ergebnisse der MRT, also zum Beispiel die Anzahl an neuen T₂-Läsionen nach 6 Monaten oder später, herangezogen, zum Anderen klinische Resultate, wie die bei den Patienten auftretende Exazerbationsrate. Aufgrund verschiedener Zielkriterien und unterschiedlicher Angaben in den einzelnen Studien war eine direkte Vergleichbarkeit der meisten Studien untereinander nicht möglich, aber immerhin konnte ein Behandlungseffekt so jeweils für das in der entsprechenden Studie untersuchte Kriterium nachgewiesen werden.

2.4.3. Anwendbarkeit

Mithilfe der aus den ermittelten validen und wichtigen Studien errechneten Werte wurde in der vorliegenden Arbeit schließlich diejenige Behandlungsmöglichkeit ermittelt, die für den

anfangs im Szenario vorgestellten Patienten die geeignetste ist, um dessen Anzahl der jährlichen Exazerbationen als klinisch wichtigstes Outcomekriterium und die Menge der Gehirnläsionen als pathologisches Korrelat der Krankheitsaktivität zu vermindern. Nähere Ausführungen zum Zusammenhang zwischen MR-Ergebnissen und klinischen Befunden finden sich in Punkt 3.2.2 dieser Arbeit.

Im Vordergrund der EBM hat immer die Prüfung der Anwendbarkeit für den Patienten, für den die Literatursuche durchgeführt wurde, zu stehen. Dabei sollten die Einschlusskriterien der Studienpatienten Ähnlichkeit mit den Daten des eigenen Patienten haben, denn je größer die Abweichung, desto höher ist die Wahrscheinlichkeit, dass die Ergebnisse der Studie für diesen nicht gelten. Man sollte die gewonnenen Erkenntnisse verwerfen, wenn man aus seiner klinischen Erfahrung heraus urteilt, dass sich die Kriterien des Patienten von denen der Studienteilnehmer zu sehr unterscheiden und man nicht mehr davon ausgehen kann, dass die Ergebnisse auch auf den eigenen Patienten zutreffen (Sackett 1999, 129f. [134]).

Außerdem muss geprüft werden, ob der Patient einer Untergruppe der untersuchten Studienpopulation angehört und ob die Ergebnisse für diese Subgruppe von den Hauptergebnissen abweichen. In der Regel kann man annehmen, dass die für die Gesamtpopulation einer Studie gefundenen Ergebnisse auch auf Untergruppen anwendbar sind (Sackett 1999, S.130 [134]). Falls aber die Patienten der Untergruppe von der geprüften Therapie im Gegensatz zur Gesamtpopulation einer Studie nicht profitieren und der eigene Patient tatsächlich seinen Merkmalen nach dieser Untergruppe angehört, müssen weitere Fragen beantwortet werden. Man kann die Ergebnisse der Studie verwerten, wenn man alle folgenden Fragen mit „Ja“ beantworten kann (Sackett 1999, S.130 [134]) :

- 1) Ist der Unterschied wirklich biologisch und klinisch plausibel?
- 2) Ist der qualitative Unterschied sowohl klinisch (in einigen Fällen nützlich, in anderen Fällen dagegen nutzlos oder sogar schädlich) als auch statistisch signifikant?
- 3) Wurde der Unterschied schon vor der Studie als Hypothese formuliert (oder ergibt er sich lediglich aus dem intensiven „Aussaften“ der Daten), und wurde er in anderen, unabhängigen Studien bestätigt?
- 4) Handelt es sich um eine von nur wenigen Untergruppenanalysen, die in dieser Studie durchgeführt wurden?

Falls eine dieser Fragen mit „Nein“ beantwortet wird, bestehen berechtigte Zweifel an der Validität und damit der Aussage der Subgruppenanalyse. Sollte man jedoch aufgrund seiner klinischen Erfahrung der Meinung sein, dass die Therapie für einen bestimmten Patienten nützlich ist, obwohl er einer Subgruppe mit anderen Ergebnissen angehört, folgt als nächster Schritt die Frage, wie groß der mögliche Nutzen für diesen Patienten ist. Man kann auch zum nächsten Schritt der EBM gehen, wenn der Patient keiner Untergruppe, sondern der Gesamtpopulation angehört oder wenn die vier genannten Fragen mit „Ja“ beantwortet wurden.

Der NNT-Wert aus der Studie muss also in einen für den Patienten individuellen NNT-Wert umgerechnet werden. Dazu gibt es zwei Möglichkeiten. Die erste Möglichkeit passt mit Hilfe einer Literaturrecherche zur Anfälligkeit des individuellen Patienten den NNT-Wert aus der Studie an. Falls die Eigenschaften des Patienten mit denen der Studienteilnehmer übereinstimmen, kann das Ergebnis für seine Anfälligkeit aus der Therapiestudie ohne zusätzliche Literatursuche direkt übernommen werden. Andernfalls kann eine Recherche durchgeführt werden, um Informationen über die Prognose des Patienten ohne Behandlung seiner Erkrankung zu erhalten. Mit dem gefundenen Ergebnis, das die bei ihm wahrscheinlich auftretende Ereignisrate (patient expected event rate = PEER) in Prozent oder als Dezimalzahl widerspiegelt, lässt sich die NNT für den Patienten errechnen. Dazu geht man wie folgt vor: Man nimmt an, der aus der Studie erhaltene Wert für die RRR sei unabhängig vom Ausgangsrisiko. Dies ist zwar nicht ganz korrekt, für die klinischen Belange in der Regel aber ausreichend. Dann bestimmt man die NNT wie folgt: $NNT = 1/ARR = 1/(PEER * RRR)$. (Sackett 1999, S.131 [134]) Falls sich aus einer Literaturrecherche zum Beispiel eine PEER von 50% oder 0,5 für ein negatives Ereignis ergibt und die aus der Therapiestudie errechnete RRR 60% oder 0,6 beträgt, ergäbe sich die NNT dann nach oben stehender Formel: $NNT = 1/(0,5 * 0,6) = 3,33$. Es müssten in diesem Fall also vier Patienten behandelt werden, um bei einem das Auftreten des unerwünschten Ereignisses zu verhindern. Alternativ kann das Ergebnis auch aus dem Nomogramm in Abbildung 8 des Anhangs abgelesen werden (Sackett 1999, S. 131ff. [134]).

Die zweite Möglichkeit zur Anpassung der NNT geht von einer Schätzung aufgrund der klinischen Erfahrung des behandelnden Arztes aus. Dieser muss dazu die Anfälligkeit seines Patienten im Vergleich zur Studienpopulation beurteilen und die aus der Studie erhaltene NNT dann durch das Schätzergebnis dividieren (Sackett 1999, S.132 [134]). Falls die NNT aus der Studie zum Beispiel 5 ergäbe und das Risiko, ein unerwünschtes Ereignis zu erhalten, auf 0,5, also die Hälfte der Anfälligkeit der Studienteilnehmer, geschätzt würde, ergäbe sich dann folgende NNT: $NNT = 5/0,5 = 10$. Die Anzahl Patienten, die behandelt werden müsste, um bei einem das Auftreten eines unerwünschten Ereignisses zu verhindern, wäre in diesem Beispiel also 10.

Nach Erhalt einer individuellen NNT müssen schließlich die Wertvorstellungen und Präferenzen des Patienten mit der möglichen Therapie, deren Wirkungen und Nebenwirkungen verglichen werden. Das Ergebnis dieses Vergleichs ist abhängig von dem erwarteten Nutzen der Behandlung und den dadurch entstehenden möglichen Schäden. Überwiegt der Nutzen, wird sich ein Patient schnell für die Therapie entscheiden, überwiegen die Nebenwirkungen, ist die Behandlung in der Regel unattraktiv. Oftmals existieren jedoch zwei oder mehrere gleichwertige Therapieoptionen nebeneinander oder es gibt andere triftige Gründe, einer Behandlung nicht ohne weiteres zuzustimmen. Dann dauert es meist längere Zeit, bis sich

ein Patient für oder gegen eine Therapie entschieden hat. In einem derartigen Fall kann die Durchführung einer klinischen Entscheidungsanalyse sinnvoll sein (Sackett 1999, S.133 [134]).

Da es sich in der vorliegenden Arbeit um einen fiktiven Patienten handelt, konnte auf eine Entscheidungsanalyse verzichtet und die RRR für die Bestimmung des individuellen Risikos als 1 angesehen werden. Unter diesen Vorgaben entspricht die NNT für unseren Patienten den NNTs aus der Wichtigkeitsanalyse in Abschnitt 3.1.2.2. und 3.2 dieser Arbeit.

2.5. Erstellung eines objektiveren Bewertungssystems für die Validität von Therapiestudien

Das EBM-Bewertungsschema nach Sackett bildet die Grundlage für eine Bewertung der eingeschlossenen Therapiestudien (Sackett 1999, Ss. 72-76, 105-111, 129-138 [134]). Während für die Bewertung der Wichtigkeit und die Überprüfung der Anwendbarkeit mit den Kennzahlen ARR, RRR und NNT objektive Kriterien zur Verfügung stehen, basierte bislang die Beurteilung der Validität nicht auf derart harten Fakten. Vielmehr sind manche Fragen oftmals nicht eindeutig zu beantworten, da in den bewerteten Studien die entsprechenden Angaben fehlen oder unvollständig sind. Deshalb wurde in der Validitätsbeurteilung der Studien ein auf dem Sackettschen Fragebogen basierendes Bewertungssystem eingeführt, mit dessen Hilfe die Bewertung nach EBM objektiviert werden soll und das im Folgenden „System nach Hassler“ genannt wird. Pro Schritt werden je nach Gewichtung zwischen fünf und sieben Teilpunkte vergeben, und es kommt zum Punktabzug, wenn ein Merkmal nicht erfüllt oder nicht ausreichend im Bericht der Studie beschrieben wurde. Am Ende steht aufgrund der Gesamtzahl an Punkten und anderen Kriterien die Beurteilung einer Studie als voll, eingeschränkt oder nicht valide in fünf Stufen von 4 bis 0 (Abbildung 9 im Anhang). Ziel dieses Systems ist also die Ermöglichung einer objektiveren Bewertung von Therapiestudien, die dann hinsichtlich ihrer Validität direkt miteinander verglichen werden können. Dabei erfolgt das Vorgehen für die Punktvergabe in den einzelnen Schritten nach bestimmten Kriterien, die ab Seite 36 in den jeweiligen Abschnitten genauer beschrieben werden. Bei Vorhandensein eines Merkmals kann diesem in vielen Fällen eine eindeutige Punktzahl zugewiesen werden und damit die Beurteilung objektiver erfolgen. In Situationen, in denen der Bewerter unsicher ist, gibt es die Möglichkeit, halbe Punkte zu vergeben, bei komplett fehlenden Angaben wird der entsprechende Schritt mit null Punkten bewertet.

Wichtig ist auch zu beachten, dass bei =2 Teilpunkten in einem der Schritte, unabhängig von der Gesamtpunktzahl, eine Abwertung der untersuchten Studie auf maximal „eingeschränkt valide“ mit Validitätsstufe 2 erfolgt und bei 0 Punkten in einem der Schritte die Studie auf

„nicht valide“, das heißt Validitätsstufe 0, herabgestuft wird. Die einzelnen Kriterien des Systems nach Hassler sind so ausgelegt, dass zwei Punkte oder weniger dann vergeben werden, wenn die RCT aufgrund einer niedrigeren Qualitätsstufe keinesfalls mehr voll valide sein kann. 0 Punkte in einem der Einzelschritte bedeutet, dass dessen Forderungen überhaupt nicht berücksichtigt wurden und somit ein ganz essenzielles Kriterium für die Validität fehlt. Dann kann eine Studie nur noch als „nicht valide“ eingestuft werden. Als Ausnahme führt das Fehlen einer Randomisierung (Schritt 1), selbst wenn die Gesamtpunktzahl im Schritt 1 größer Null ist, wegen der Wichtigkeit dieser Forderung trotzdem zur Abwertung des entsprechenden Schrittes auf null Punkte und damit der gesamten Studie auf „nicht valide“. Das gleiche gilt für eine ungenügende Verblindung (Schritt 4) sowohl von behandelnden oder diagnostizierenden Ärzten als auch Patienten in zwei der drei Fälle, das heißt, wenn im Schritt 4 zweimal null Punkte vergeben werden müssen. Somit wird der großen Wichtigkeit von Randomisierung und Verblindung nochmals verstärkt Rechnung getragen.

Für die Bewertung der Studien in der vorliegenden Arbeit wurde zunächst eine klassische Auswertung der eingeschlossenen 38 MS-Therapiestudien nach Sackett (Sackett 1999, Ss. 72-76, 105-111, 129-138 [134]) vorgenommen. Anschließend erfolgte eine erste Durchsicht der Studien mit dem weiterentwickelten System und anhand der Ergebnisse ein Abgleich mit den Resultaten des herkömmlichen Schemas. Danach wurden die Punktvergabe für die einzelnen Schritte und die Beurteilungsstufen der Validität bestimmt. Schließlich wurden in einem weiteren Durchgang die festgelegten Abschnitte und die Punktvergabe validiert, wo nötig nochmals verändert und abschließend die endgültigen Bewertungsergebnisse ermittelt, sodass die Einschätzung der Validität mithilfe des Systems nach Hassler derjenigen der äußerst sorgfältig durchgeführten Bewertung nach Sackett entsprach. Das System nach Hassler wurde also im Vergleich mit der Vorlage nach Sackett evaluiert und optimiert, sodass mit der in Abbildung 9 gezeigten Version ein objektiveres Werkzeug zur Beurteilung der Validität von Therapiestudien aller Art zur Verfügung steht.

2.5.1. Randomisierung

Im ersten Schritt des Systems nach Hassler, der nach der Zufallszuteilung, deren Zeitpunkt und deren Geheimhaltung fragt, werden für die Randomisierung drei Punkte vergeben, wenn diese sicher stattfand und valide durchgeführt wurde, zwei Punkte, wenn eine Randomisierung zwar angegeben ist, aber dabei kleinere Fehler begangen worden sind, und ein Punkt, wenn eine Randomisierung zwar angegeben ist, dabei jedoch ein größerer oder mehrere kleinere Fehler gemacht wurden. Null Punkte sind zu vergeben, wenn in der Studie keine Zufallszuteilung stattfand. Sollte keine Randomisierung stattgefunden haben, ist Schritt 1 auf

null Punkte abzuwerten und führt damit insgesamt zu einer Abwertung auf Validitätsstufe 0 („nicht valide“), denn nicht randomisierte Studien sollten für Therapieentscheidungen nur im Notfall herangezogen werden, das heißt, wenn es mit Sicherheit keine anderen Quellen für Evidence zu der Behandlung einer Krankheit gibt (Sackett 1999, S.74 [134]).

Wichtig ist jedoch nicht nur die Frage nach der Randomisation, sondern auch, ob sie zum richtigen Zeitpunkt durchgeführt wurde. Falls zum Beispiel eine Zufallszuteilung schon zu Beginn einer Studie stattfindet, anschließend aber aufgrund einer Änderung der Einschlusskriterien oder wegen der Notwendigkeit einer nicht mit der Studienmedikation kompatiblen Therapie bestimmte Patienten doch nicht für die entsprechende Studie in Betracht kommen, wird das Ergebnis verfälscht werden. Deshalb ist der Zeitpunkt der Randomisation, obwohl im Schema nach Sackett nicht ausdrücklich erwähnt, essenziell für die Validität. Zwei Punkte sind für die vollständig richtige Terminierung, ein Punkt bei kleineren Fehlern und null Punkte für eine zeitlich falsch gelegte Randomisierung zu vergeben.

Außerdem ist schließlich noch, wie auch bereits in Gliederungspunkt 2.2.1.1. beschrieben, die Geheimhaltung der Zufallszuteilung wichtig. Damit beschäftigt sich die dritte Frage im ersten Abschnitt. Im Unterschied zur Verblindung geht es dabei ausschließlich um den Verschluss der die Randomisierung betreffenden Daten. So kann eine Verblindung zum Beispiel aufgrund von Nebenwirkungen einer Therapie aufgehoben sein, obwohl die Zufallszuteilung geheimgehalten war. Umgekehrt kann es sein, dass eine Randomisierung und damit auch eine Geheimhaltung gar nicht stattfand, zum Beispiel wenn Patienten nach bestimmten Kriterien einer Behandlung zugewiesen werden, die nur einer externen Zuteilungsstelle bekannt sind. In diesem Fall könnten aber trotzdem Patienten und ein Teil der beteiligten Ärzte im Unklaren über die zugewiesenen Therapieschemata sein. Für die Geheimhaltung sind zwei Punkte zu vergeben, wenn sie aus dem Studienreport mit Sicherheit hervorgeht. Einen Punkt sollte man in zwei Fällen zuteilen: Entweder, wenn eine Geheimhaltung zwar in der Studie angegeben, aber nicht komplett eingehalten worden war, oder wenn eine Geheimhaltung nicht angegeben, aber aus den sonstigen Studiendaten zu entnehmen ist, dass keine der an der Studie beteiligten Personen von der Verteilung wusste. Sollte eine Versiegelung der Zufallszuteilung in der bewerteten Studie nicht stattgefunden haben, sind null Punkte zu vergeben.

2.5.2. Berücksichtigung aller eingeschlossenen Patienten in der Gesamtwertung

Schritt zwei der objektiven Validitätsbewertung einer Therapiestudie überprüft, ob alle zu Beginn einer Studie eingeschlossenen Patienten am Ende ausgewertet wurden. Zur Errechnung der Prozentzahl ist die am Ende der Studie einbezogene Zahl an Patienten durch die

Anzahl der anfangs in die Studie eingeschlossenen Patienten zu dividieren. Dabei werden sechs Punkte vergeben, wenn mehr als 95% aller Patienten am Ende zur Auswertung gelangen. Die Punktaufteilung ist in 5%-Schritte gestaffelt bis zu null Punkten bei am Ende weniger als 70% in die Auswertung eingeschlossenen Patienten. Erwähnenswert ist die 80%-Grenze, bei deren Überschreiten nach oben 3 Punkte zugeteilt werden. Wie in Abschnitt 2.5 dieser Arbeit beschrieben, existieren Grenzwerte, ab denen eine Studie bestimmte Validitätsstufen nicht mehr erreichen kann. Im Falle der Frage nach Berücksichtigung von Patienten in der Endauswertung kann eine RCT nur dann unter Umständen noch voll valide sein, wenn mindestens 80% aller zu Beginn eingeschlossenen Patienten am Ende ausgewertet werden (Sackett 1999, S.75 [134]). Wie in Punkt 2.2.1.2. beschrieben, sollte man bei allen verlorengegangenen Patienten den schlimmstmöglichen Fall, das sogenannte „Worst-case-scenario“, annehmen können. Bei weniger als 80% berücksichtigten Patienten ist dies meistens nicht mehr möglich (Sackett 1999, S.75 [134]), und eine derartige Studie kann nur noch maximal eingeschränkt valide mit Stufe 2 sein.

2.5.3. Intention-to-treat-Analyse

Das nächste für die Validität wichtige Kriterium ist die „Intention-to-treat“-Analyse (ITT). Wie in Abschnitt 2.4.1.3. beschrieben, ist es wichtig, dass die Patienten in den Gruppen ausgewertet werden, denen sie zugeordnet waren. Bei der objektiven Bewertung nach Hassler sind vier Punkte zu vergeben, wenn eine ITT sicher durchgeführt wurde, drei Punkte, wenn eine ITT mit großer Wahrscheinlichkeit stattfand, einer, falls man nach Prüfung der Datenlage zu dem Entschluss kommt, eine ITT sei eher nicht durchgeführt worden und null Punkte für ein sicheres Fehlen einer ITT oder bei völliger Unsicherheit aufgrund fehlender Angaben, sodass auch aus allen anderen Daten der bewerteten Studie nicht ersichtlich ist, ob eine ITT stattfand oder nicht. Sollte der Bewerter sich nicht für eine der beiden Alternativen ein oder drei Punkte entscheiden können, werden in Ausnahmefällen zwei Punkte vergeben. Um dem vorzubeugen, ist diese Alternative jedoch auf dem Bewertungsbogen nicht vermerkt.

Als weiterer Bewertungsaspekt ist ein Zusatzpunkt zu vergeben, wenn die Auswertung nach ITT erwähnt wird, da diese Angabe als Hinweis zu werten ist, dass sich die Autoren Gedanken zu diesem Thema gemacht haben, was als positiv zu gelten hat. Allerdings ist diese Tatsache nicht überzubewerten, weshalb dafür 1 Punkt vergeben wird, der nicht gegen eine Abwertung benutzt werden kann. Das bedeutet, dass dieser Punkt zwar in die Gesamtzahl am Ende eingerechnet wird, falls aber trotz Erwähnung einer ITT diese sicher *nicht* durchgeführt wurde, muss der Schritt 3 zur Abwertung der Studie auf Validitätsstufe 0 („nicht valide“) führen. Wichtig ist nicht die Angabe über eine ITT, sondern ob sie wirklich stattfand.

2.5.4. Verblindung

Auch für die Verblindung gilt, dass es eher unwichtig ist, ob sie im Text einer RCT Erwähnung findet, sondern vielmehr, dass sie wirklich praktiziert wurde. Zu ihrer Beurteilung folgt Schritt vier, der in drei Fragen unterteilt ist: Die erste bewertet die Verblindung des Patienten, die zweite behandelt die Verblindung des Therapeuten und die dritte die des diagnostizierenden Arztes. Zwei Punkte sind jeweils zu vergeben, wenn man die Frage mit Sicherheit mit „Ja“ beantworten kann, ein Punkt, wenn nicht eindeutig sichergestellt ist, ob eine Verblindung exakt eingehalten wurde, aber keine Hinweise vorliegen, dass sie fehlte, jedoch auch, wenn eine Verblindung ursprünglich gegeben war, diese aber, zum Beispiel aufgrund von Nebenwirkungen einer Behandlung, zum Teil aufgehoben war. Null Punkte sollten eingetragen werden, wenn die Verblindung fast vollständig aufgehoben oder wenn sie von vornherein nicht vorhanden war.

Die Regel, dass eine Abwertung der Studie auf „nicht valide“ (Stufe 0) erfolgt, wenn ein Teilschritt mit null Punkten bewertet wurde, kommt hier im Schritt 4 des Systems nach Hassler schon dann zum Tragen, wenn zwei der drei Fragen mit „Nein“ beantwortet werden müssen. In diesem Fall kann nicht mehr mit rein objektiven Ergebnissen einer Studie gerechnet werden, und aufgrund einer niedrigeren Qualitätsstufe muss die entsprechende Studie dann mit „nicht valide“ bewertet werden. Es ist an dieser Stelle jedoch, wie schon im Abschnitt 2.4.1.4. dieser Arbeit, noch einmal zu betonen, dass Studien, die die Bewertung Validitätsstufe 0 aufgrund Abwertung im Schritt 4 erhalten, zwar nicht den strengen Anforderungen der EBM entsprechen, in Fällen, in denen jedoch keine andere Evidence zu der untersuchten Therapie in der Literatur zu finden ist, trotzdem unter Umständen ihren Stellenwert haben können.

2.5.5. Gleichbehandlung

Im fünften Schritt des objektiveren Bewertungsschemas nach Hassler folgt die Bewertung, ob die Vergleichsgruppen außer der Studienmedikation ansonsten identisch behandelt wurden. Wie schon im Gliederungspunkt 2.4.1.5. beschrieben, ist dies für die Beurteilung der Validität einer Studie ein weiteres wichtiges Kriterium. Bei der Bewertung sind maximal fünf Punkte zu vergeben, und zwar dann, wenn der Leser mit Sicherheit feststellen kann, dass eine Gleichbehandlung stattfand. Falls eine identische Behandlung wahrscheinlich ist, werden vier Punkte, und falls sie eher gegeben, nicht eindeutig aus den Studiendaten zu eruieren ist, aufgrund des Berichts der Studie aber nicht unwahrscheinlich ist, drei Punkte zugeteilt. Auch wenn die Vergabe von zwei Punkten durchaus erlaubt ist, wird im Schritt fünf mit Absicht, wie im Bewertungsschritt „ITT“, diese Stufe im Bewertungsbogen ausgelassen. Da-

mit soll der Unterschied zwischen einer eher gegebenen und einer eher nicht erfolgten Gleichbehandlung, für die es einen Punkt gibt, zum Ausdruck kommen. Keinen Punkt gibt es für eine mit Sicherheit nicht erfolgte Gleichbehandlung.

2.5.6. Ähnlichkeit der Patienten

Bei der Überprüfung der Ähnlichkeit muss die Beurteilung etwas subjektiver erfolgen. Trotz aller Bemühungen, eine Bewertung auf diesem Feld zu objektivieren, ist es dem Beurteiler überlassen, einzelne Kriterien mithilfe seines medizinischen Wissens zu gewichten und auszuzählen. Die Ähnlichkeit der Patienten in den Vergleichsgruppen ist essenziell für die Aussagekraft einer Studie, denn falls Unterschiede schon im Voraus bestehen, wird auch am Ende einer Studie unklar bleiben, ob ein potentieller Therapieeffekt den Unterschieden zu Beginn oder der eigentlichen Behandlung zuzuschreiben ist. Die einzige Ausnahme hiervon sind Studien, die explizit darauf abzielen, einen Vergleich mit gesunden Probanden, die zwangsläufig Unterschiede zu den Studienpatienten aufweisen müssen, zu unternehmen.

Bei der Beurteilung der Ähnlichkeit geht man wie folgt vor. Zunächst überprüft man, ob die Studienpatienten alle an der gleichen Krankheit, deren Therapieoptionen verglichen werden sollen, leiden. Sollte es dabei keine Übereinstimmung geben, was allerdings nur äußerst selten der Fall ist, sind in diesem Schritt ohne weitere Überprüfung null Punkte zu vergeben und damit die Studie auf „nicht valide“ herabzustufen. In der Regel lässt sich diese Frage jedoch mit „ja“ beantworten.

Als nächstes sucht man sich mithilfe seiner internen Evidence, das heißt seines eigenen medizinischen Wissens, oder über Sekundärliteratur die Eigenschaften der Studienpatienten, die in der untersuchten Studie für das Outcome besonders relevant sind, kennzeichnet diese in der vorderen Spalte mit einem „X“ oder „XX“ und wichtet diese in der Punktvergabe für Schritt 6 einfach oder doppelt. Für die unwichtigen Kriterien trägt man eine „0“ ein und kann Differenzen der Behandlungsgruppen in diesen Eigenschaften eher vernachlässigen. Außerdem sucht man sich neben den bereits aufgeführten Kriterien weitere, die für die entsprechende Therapie wichtig sein könnten, trägt diese am Ende der Tabelle ein und versieht sie je nach Gewichtung mit einem „X“ oder „XX“.

In vielen Studien wird man die nötigen Daten in einer Tabelle, meist „table 1“, aufgelistet vorfinden, wobei es in Multicenter-Studien besonders gut ist, wenn zusätzlich eine Aufschlüsselung der Patientendaten nach den an der Studie teilnehmenden Zentren erfolgt. Sollte eine Darstellung in einer Tabelle nicht aufgeführt sein, können manche Eigenschaften auch aus den Ein- und Ausschlusskriterien oder anderen Textstellen entnommen werden. Eine Beurteilung alleine anhand der Kriterien, die die Patienten für die Teilnahme an einer Studie erfül-

len müssen, kann jedoch auch problematisch sein, da mit der Nennung dieser Eigenschaften noch nicht sichergestellt ist, dass diese auch wirklich in ihrem Schweregrad auf die Gruppen gleich verteilt sind.

In der Bewertung der Ähnlichkeit sind für die Übereinstimmung der Vergleichskollektive insgesamt maximal sechs Punkte zu vergeben. Abhängig von dem Grad der Übereinstimmung werden dafür zwischen 0 und fünf Punkte für die Ähnlichkeit der Patienten und zusätzlich ein Punkt für das Vorhandensein einer Tabelle mit den relevanten Patientendaten zugeteilt. Dieser Punkt wird deshalb vergeben, weil Autoren mit einer derartigen Tabelle zeigen, dass ihnen Transparenz in Bezug auf die Vergleichbarkeit der Gruppen wichtig ist und weil sie damit dem Leser eine Beurteilung wesentlich erleichtern. Der Zusatzpunkt für das Vorhandensein einer Table 1 ist jedoch in Klammern aufgeführt, weil er, ebenso wie der zusätzliche Punkt im Schritt 3 (ITT), zur Gesamtpunktzahl dazugezählt wird, aber nicht verhindern kann, dass, wenn die Studienteilnehmer der Behandlungsgruppen nicht vergleichbar sind (also 0 Punkte vergeben werden), eine Abwertung der Studie auf Validitätsstufe 0 erfolgt.

2.5.7. Vorgehen bei der Zusammenfassung der Ergebnisse aus der objektiven Bewertung

Mit der Bewertung der Übereinstimmung der Vergleichsgruppen ist der erste Teil der Validitätsprüfung abgeschlossen. Es folgen die Schritte **A**, **B**, **C** und **D**, in denen die bisherigen Ergebnisse zusammengefasst und ergänzt werden.

Im Schritt **A** erfolgt ein Übertrag der Einzelpunkte aus den Schritten 1 bis 6 und eine vorläufige Bewertung der untersuchten Studie nach dem Schema in Abbildung 10.

Gesamtbewertung:		
Voll valide (4)	31 – 35	Punkte
Eingeschränkt valide (3)	28 – 30	Punkte
Eingeschränkt valide (2)	24 – 27	Punkte
Eingeschränkt valide (1)	19 – 23	Punkte
Nicht valide (0)	0 – 18	Punkte

Abbildung 10: Bewertung einer Studie abhängig von der im objektiven Bewertungsschema erreichten Punktzahl (siehe auch Tabelle 6).

Im Schritt **B** wird anschließend geprüft, ob in einem der Schritte 1 bis 6 ein Grund vorliegt für eine mögliche Abwertung der Studie auf Validitätsstufe 2, falls =2 Punkte in einem der Schritte vergeben wurden, oder auf Validitätsstufe 0, falls in einem der Schritte 0 Punkte vergeben wurden.

Nach Eintragung der dann maximal möglichen Bewertung geht man über zum Bewertungsschritt **C**, in dem untersucht wird, ob es weitere Anhaltspunkte für eine Einschränkung der

Validität geben könnte. Dies ist der subjektivste aller Schritte und basiert auf dem Vorschlag von Porzsolt und Kollegen, die zusätzlich zu den herkömmlichen sechs Schritten in Sacketts EBM-Bewertung die Einführung eines siebten Schrittes fordern, mit dessen Hilfe man alle Aspekte, die in den Schritten eins bis sechs nicht beachtet wurden, in die Bewertung einfließen lassen kann (Porzsolt 2001, S.26 [125]). Hier muss der Bewerter allein aufgrund seiner Erfahrung und seines in seinem Fach erworbenen Wissens um die Wichtigkeit zusätzlicher Schwierigkeiten in einer Studie bewerten, wie hoch die Validität nach Prüfung weiterer Anhaltspunkte für Fehler der untersuchten Studie dadurch am Ende noch maximal sein kann.

Im Schritt **D** bildet man schließlich eine Synthese, indem man die maximal möglichen Bewertungen in **A**, **B** und **C** überprüft und daraus das Endergebnis der Bewertung ableitet.

2.6. Vorgehen bei der Erstellung eines Reviews zur Therapie der Multiplen Sklerose

Nach Abschluss der Bewertung der in die vorliegende Arbeit eingeschlossenen Therapiestudien wurden deren Resultate mit denen eines bereits bestehenden systematischen Reviews verglichen und daraus Empfehlungen für die Therapie bei MS abgeleitet. Der Begriff „systematisches Review“ bedeutet eine gezielte, auf *Vollständigkeit* ausgelegte Suche nach relevanten Studien, deren Auswahl und kritische Bewertung. Die Ergebnisse der verwendeten Publikationen werden analysiert, ausgewertet und zusammengefasst (www.cochrane.org [166]). Da das primäre Ziel der vorliegenden Arbeit ausdrücklich nicht eine vollständige Zusammenfassung der die Therapie bei MS behandelnden Studien war, muss in diesem Fall lediglich von einem sogenannten unsystematischen oder einfachen Review statt eines *systematischen* Reviews gesprochen werden.

Ziel der Empfehlungen in der vorliegenden Arbeit ist, im Gegensatz zur Therapie akuter Exazerbationen oder chronischer Symptome, die Langzeitbehandlung, also der Versuch der Hemmung von Krankheitsprogression und Rückfallraten. Für die Suche nach einer Metaanalyse für die Langzeittherapie der MS wurden die Internetseiten der Cochrane-Collaboration durchsucht und als einziges bestehendes Review mit einer Empfehlung der Collaboration folgender Artikel gefunden, der zum Vergleich mit den Ergebnissen aus der vorliegenden Arbeit diente: „Disease-modifying drugs for multiple sclerosis: a rapid and systematic review“ (Clegg 2000 [67]). Unterstützend wurden die Nachschlagewerke „Pathogenese und Therapie der Multiplen Sklerose“ (Gold 2000 [84]) und „Therapie der Multiplen Sklerose“ (Wötzel 2000 [158]) herangezogen. Für das Review der vorliegenden Arbeit kamen nur die Studien zum Einsatz, die in der Validitätsbeurteilung mindestens Validitätsstufe 2 nach dem objektiven Bewertungsschema erhalten hatten und aus deren Ergebnissen die Kennzahlen der Wichtigkeit RRR, ARR und NNT zu errechnen waren. Mehrfachpublikationen zu einer Studie wären

zusammengenommen als eine Studie betrachtet worden, um Berechnungen von Patienten nicht doppelt zu werten (siehe Tabelle 4 in Abschnitt 3.1.1.). Allerdings waren unter den 14 am Ende übriggebliebenen Studien keine, die in Form einer Zweitpublikation häufiger als einmal in die Auswertung gelangten. Das untenstehende Diagramm 1 gibt die Auswahl der Veröffentlichungen für das unsystematische Review der Therapieoptionen bei MS wieder (siehe Abbildung 2).

Um den Stellenwert der MRT für Therapiestudien besser abschätzen zu können, wurde außerdem aus Reviews und Einzelstudien die Bedeutung der MRT bei MS zusammengestellt und ausgewertet.

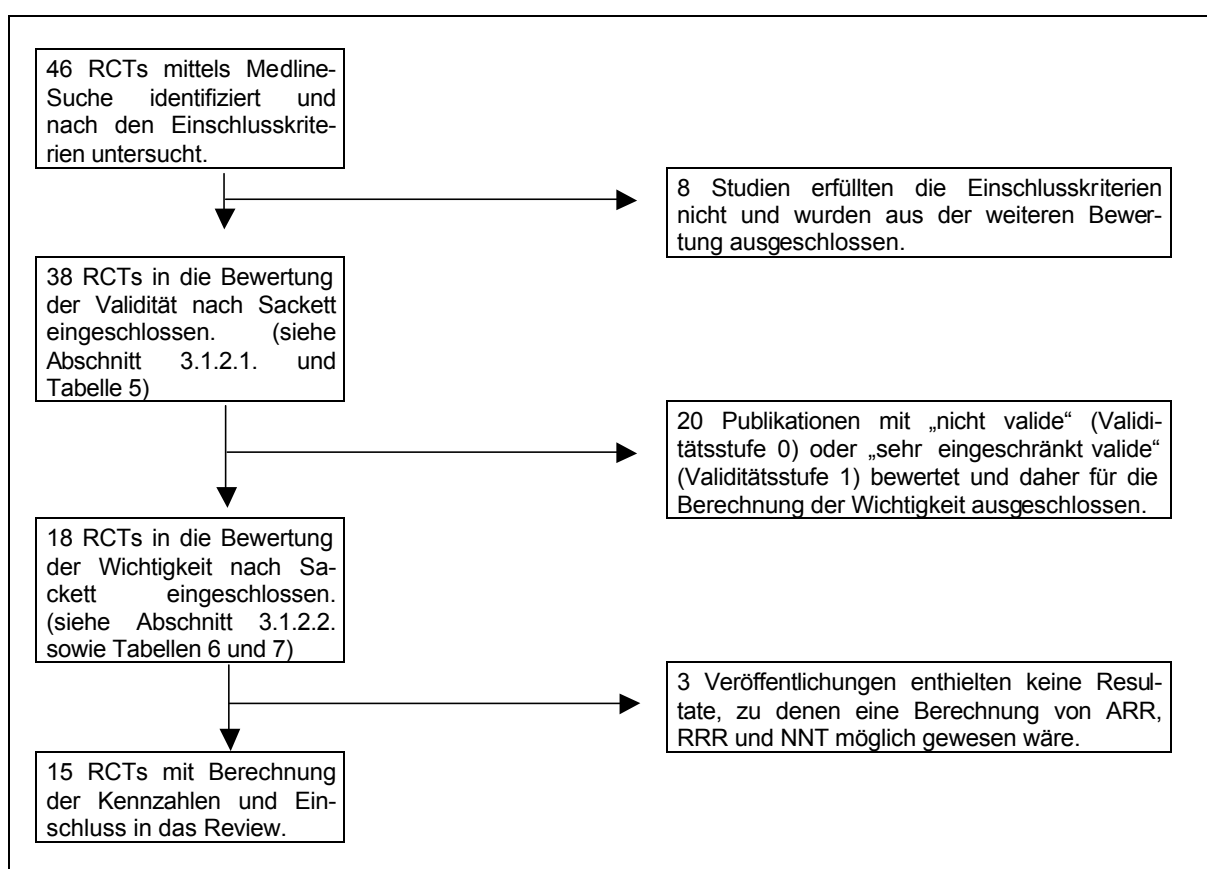


Diagramm 1: Auswahlsschritte für den Einschluss von Veröffentlichungen in das sogenannte einfache oder unsystematische Review dieser Arbeit.

3. Ergebnisse

Im folgenden Abschnitt dieser Arbeit werden zunächst die Ergebnisse der Auswertung und Evaluation der eingeschlossenen Studien beschrieben, um anschließend in kompakterer Form die Resultate eines unsystematischen Reviews der bewerteten Therapiestudien mit derzeitigen Behandlungsoptionen und der Bedeutung der MRT bei MS vorzustellen.

3.1. Ergebnisse der Untersuchung von Validität, Wichtigkeit und Anwendbarkeit der Therapiestudien

Hier werden die Ergebnisse der Bewertung vorgestellt und positive Aspekte sowie häufige Probleme vieler Therapiestudien aufgezeigt.

Viele für eine Auswertung wichtige Fragen lassen sich oftmals mangels Daten nicht eindeutig beantworten. In diesen Fällen wurde, wo sinnvoll, versucht, aus anderen Informationsquellen, wie zum Beispiel aus Studien, die als Fußnote in der bewerteten Publikation genannt wurden (=Schwesterstudien) oder durch Untersuchung anderer als der üblichen Abschnitte einer Veröffentlichung, die nötigen Daten zu erhalten. Dies geschah zum Beispiel durch Suche nach Informationen im Ergebnisteil, obwohl man diese eigentlich unter „Material und Methoden“ zu finden hätte.

Diese Arbeit soll keine Kritik an weniger gut durchgeführten Studien vornehmen, sondern gute Beispiele hervorheben, allgemeine Fehler aufdecken und daraus Verbesserungsvorschläge ableiten. Aus Gründen der Nachvollziehbarkeit kann jedoch bei einem direkten Verweis auf diese Studien oder bei wörtlich übernommenen Aussagen nicht auf eine Zitation von fehlerhaften Literaturstellen verzichtet werden. Um die Lesbarkeit zu erhalten, sind allgemeine Aussagen über das Gesamtkollektiv untersuchter Publikationen nicht mit Fußnoten versehen. Die nötigen Informationen können im Anhang aus den Tabellen 8 (CONSORT) und 5 (EBM) ersehen werden, in denen die Resultate aufgelistet sind.

3.1.1. Auswertung des Fragebogens für zusätzliche Aspekte der verwendeten Studien

Von den 38 in die Bewertung eingeschlossenen Berichterstattungen über klinische Studien lagen dreizehn vor, die zwei- oder sogar dreimal in Form von Sekundärberichterstattungen oder Subgruppenanalysen über die durchgeführten Studien berichteten. Weil damit über die selbe Studie mehrfache Daten vorlagen, wurde nur jeweils einer der Berichte über diese Untersuchungen in die Statistik über die Daten aus der Gesamtstudie aufge-

Mehrfachberichte über die selbe klinische Studie in:	Thema	In Studienstatistik verwendet
<ul style="list-style-type: none"> • Filippi, Neurol Sci 2000 [8]; • Filippi, Neurology 2000 [7]; • Rice 2000 [32] 	Cladribin	Filippi, Neurol Sci 2000 [8]
<ul style="list-style-type: none"> • Sørensen, Multiple Sclerosis 2000 (Dänischer Teil) [38]; • Sørensen, Neurology 1998 [37] 	Immunglobulin G	Sørensen, Multiple Sclerosis 2000 (Dänischer Teil) [38]
<ul style="list-style-type: none"> • Jacobs, N Engl J Med 2000 [13]; • Nortvedt 1999 [26] 	Interferon β -1a	Jacobs, N Engl J Med 2000 [13]
<ul style="list-style-type: none"> • Paolillo 1999 [30]; • Koudriavtseva 1998 [17] 	Interferon β -1a	Paolillo 1999 [30]
<ul style="list-style-type: none"> • Simon, 1998 [36]; • Jacobs, J Neuroimmunol 2000 [12] 	Interferon β -1a	Simon, 1998 [36]
<ul style="list-style-type: none"> • Cook, 1997 [3]; • Rohowsky-Kochan 1997 [33] 	Steroide + LK-Bestrahlung	Cook, 1997 [3]

Tabelle 4: Verwendete Publikation bei Mehrfachberichten der selben Studie.

nommen (siehe Tabelle 4). Fragen des CONSORT-Statements, die sich auf den einzelnen Bericht über eine Studie beziehen, wurden jedoch bei jedem eingeschlossenen Studienbericht analysiert. Ebenso wurde bei den Untersuchungen mithilfe der EBM-Kriterien verfahren, da diese auch innerhalb der selben Studie, zum Beispiel aufgrund Subgruppenanalysen, ungleiche Ergebnisse liefern können.

3.1.1.1. Ergebnisse der Datensammlung über die in die Studien eingeschlossenen Patienten und den Studienablauf

Die Gesamtzahl aller Patienten in den eingeschlossenen Studien betrug 6652, davon 2282 (33%) Männer, 3862 (56%) Frauen und aufgrund fehlender Patientenbeschreibung 508 (11%) Studienteilnehmer unklaren Geschlechts. Das Durchschnittsalter betrug 37,23, das mediane Alter über alle Patienten konnte nicht berechnet werden, es wurde jedoch der Median über das in den Studien angegebene Durchschnittsalter mit 36,5 Jahren bestimmt.

Das in der Diagnostik der MS am häufigsten eingesetzte Bewertungsschema ist der sogenannte EDSS-Score. Dieser war als Messung des Schweregrads der MS zum Zeitpunkt des Studienbeginns in 26 Studien (79%) angegeben, in vier Publikationen (13%) konnte die EDSS-Bewertung der Patienten zu Beginn der Studie aus den Angaben nicht ermittelt werden und zwei RCTs (6%) untersuchten den Verlauf der Erkrankung bei ersten Symptomen einer möglichen MS, wodurch dort noch keine Angabe möglich war (siehe Tabelle 9 im Anhang). Der durchschnittliche EDSS-Score betrug zum Zeitpunkt der Aufnahme der Patienten 3,17, der Median des EDSS-Scores 3,2. Im weiteren Verlauf der

Studien wurde der EDSS-Score in Form des primären oder sekundären Outcomekriteriums neben der MRT in 17 Fällen (53%) als einziger Score verwendet. In sieben Studien (22%) kamen neben dem EDSS ein weiterer Score und in zwei Fällen (6%) zwei zusätzliche Scores zum Einsatz. In einer Studie (3%) wurden außer dem EDSS drei weitere Scores (Neurologic Rating Scale = NRS, SF-36 Health Survey und General Health Questionnaire = GHQ) verwendet. Sechsmal, und damit am zweithäufigsten nach dem EDSS-Score, kam der NRS- bzw. SNRS- Score zum Einsatz, wobei die Abkürzung SNRS für „Scripps´ Neurologic Rating Scale“ steht und mit NRS synonym verwendet wird. Fünf Studien (16%) verwendeten keine klinischen Bewertungsschemata als Outcomekriterium. Darüber, ob sich der EDSS-Score über die Studiendauer veränderte, gab es in 13 Studien (41%) keine Angabe, 16 (50%) berichteten von einer Änderung, das heißt Verbesserung oder Verschlechterung der klinischen Symptome, und in drei Studien (9%) wurde beschrieben, dass der EDSS am Ende gleich geblieben war (siehe Tabelle 9).

Zum Einschlusszeitpunkt der Patienten in die jeweiligen Studien betrug die durchschnittliche Krankheitsdauer aller Patienten 7,67 Jahre, der Median über die durchschnittliche Dauer in den einzelnen Studien 6,5 Jahre. Die Studiendauer inklusive Nachbeobachtungsphase betrug im Mittel 669,5 Tage, die mediane Studiendauer war 710 Tage. Bei nicht einheitlichen Angaben innerhalb einer Studie wurden für die Berechnung die jeweils am längsten angegebenen Werte verwendet. Eine noch nicht abgeschlossene Studie (Koch-Henriksen 2000 [16]) und ein Studienbericht, in dem keine Angaben über die Studiendauer enthalten waren (Sailer 2000 [35]), wurden aus dieser Berechnung ausgeschlossen.

16 Studien (50%) schlossen ausschließlich Patienten mit schubweiser MS (=RRMS) ein und vier Studien (13%) nur Patienten mit sekundär progressiver MS (=SPMS). Acht Studien (25%) kombinierten Patienten mit RRMS und SPMS. In zwei Studien (6%) waren die Angaben zu ungenau, um den entsprechenden Typ der MS daraus ablesen zu können (Simon 1998, S.79 [36]: „relapsing multiple sclerosis“; Cook 1997, S.173 [3]: „clinically or laboratory supported definite MS; progressive neurological disability...“), und in zwei Studien (6%) wurde zum Typ der MS keine Angabe gemacht (siehe Tabelle 9).

Wie im Abschnitt 2.1.2. dieser Arbeit bereits beschrieben, befassten sich von den bewerteten Studien acht mit Interferon β -1a, vier mit Interferon β -1b, eine mit einem Vergleich von Interferon β -1a und Interferon β -1b, eine mit natürlichem Interferon beta und zwei mit Interferon α -2a. Des Weiteren behandelten sechs Studien monoklonale Antikörpertherapie, drei Cladribin, zwei Methylprednisolon, zwei eine Kombination aus Bestrahlung und Kortikosteroiden, zwei Mitoxantron, zwei sogenannte „myelin basic proteins“ (einmal gelöstes DR2:MBP84-102(AG284) und einmal modifizierte Peptidliganden von Myelinbasenproteinen) und jeweils eine Sulfasalazin, Lenercept und Amantadin. Die entsprechen-

den Therapieschemata wurden genau eingehalten in sechs Fällen (19%), eine weitgehende Einhaltung konnte in 3 Studien gefunden werden (9%), keine Einhaltung gab es in neun Fällen (28%) und keine Angaben in 14 Studien (44%). Zusammen genommen kommen die die Aussagekraft einer Studie einschränkenden Punkte „keine Einhaltung“ und „keine Angaben“ also in fast $\frac{3}{4}$ aller Studien vor (siehe Tabelle 9).

Da eines der Haupteinschlusskriterien für die in dieser Arbeit untersuchten Studien die Verwendung von MRT zur Diagnostik der MS ist, wird hier auch die Durchführung der MRT-Anfertigung –und Auswertung genauer beschrieben. Bei der Herstellung der Magnetresonanztomographie-Bilder verwendeten neun Studien (28%) T1-gewichtete Aufnahmen vor und nach Kontrastmittelapplikation. Zehn Studien (32%) gaben an, die T1-Images nur mit Kontrastmittel (=KM) und zwei Studien (6%), diese nur ohne Anfärbung angefertigt zu haben. Keine T1-Aufnahmen gab es in sieben Studien (22%), und zwei Studien (6%) fertigten KM-T1-Bilder nur in einer Subgruppe an (siehe Tabelle 9).

T2-gewichtete Images mit einer Echozeit zwischen 90 und 120 ms und proton-density (=PD)-Bilder mit Echozeiten zwischen 25 und 60 ms wurden in 19 RCTs (59%) angefertigt. Fünf Studien (16%) gaben an, T2-ohne PD-Gewichtung angewandt zu haben und acht Studienberichte (25%) benutzten weder T2-noch PD-Gewichtung (siehe Tabelle 9).

In zwei Studien (Coles 1999 [2], Wolinsky 2000 [45]) wurde nicht deutlich, welche Form der MRT-Gewichtung verwendet worden war, nur die Anfertigung mit Kontrastmittel wurde angegeben. Eine Studie (Optic Neuritis Study Group 1997 [29]) berichtete lediglich von Aufnahmen ohne Kontrastmaterial, die Gewichtung blieb unklar (siehe Tabelle 9).

In allen Studien, in denen Kontrastmittel zur Anwendung kam, wurde Gadolinium-Diethylentriamin-Pentaessigsäure (=Gd-DTPA) verwendet. Dies war in 23 Studien (72%) der Fall. Die Dosierung war mit einer Ausnahme von 0,2mmol/kgKG (Tan 2000 [39]) jeweils 0,1mmol/kgKG. Neun Studien (28%) führten keine KM-Untersuchungen durch (siehe Tabelle 9).

Ausgewertet wurden die MRT-Aufnahmen entweder an einer zentralen Stelle (13 Studien = 40%) oder dort, wo sie angefertigt worden waren (19 Studien = 60%). Bei letzterer Möglichkeit verteilte sich die Anzahl der Ärzte und Techniker, die die Images auswerteten, wie folgt: Je ein Arzt war in fünf Studien (26%) für die Auswertung zuständig, zwei Ärzte in sechs Studien (32%), bis zu 3 Ärzte inklusive Techniker in drei Studien (16%) und keine Angaben über die Anzahl der MR-Bewerter fanden sich in fünf Studien (26%). Von den zentral ausgewerteten Studien fand sich ein alleiniger Bewerter in 4 Fällen (33%), zwei Bewerter waren in zwei Studien (17%) zuständig und bis zu drei in einer Studie (8%). Keine Angaben über die Zahl der zentralen Bewerter fanden sich in 5 Studien (42%) (siehe Tabelle 9).

Klinische Untersuchungen fanden als primäres Zielkriterium oder neben der MRT in den meisten Studien statt: 29 RCTs (91%) haben diese Art der Auswertung zumindest erwähnt. In zwei Veröffentlichungen (6%) fanden sich keine Angaben und in einer Studie (3%) wurden sicher keine klinischen Untersuchungen durchgeführt (siehe Tabelle 9).

3.1.1.2. Ergebnisse der Untersuchung nach dem CONSORT-Statement

Hier wird dargestellt, wie häufig die einzelnen Forderungen des CONSORT-Statements erfüllt werden. Die Prozentangaben beziehen sich, wenn nicht anders beschrieben, auf die in die Auswertung eingeschlossenen 38 Publikationen.

Auch in diesem Abschnitt lassen sich einige Fragen oftmals nicht eindeutig beantworten. Manche Forderungen sind nur teilweise erfüllt, indem zum Beispiel Ergebnisse nur für die Gesamtpopulation, aber nicht für Subgruppen berichtet werden oder wenn an verschiedenen Stellen einer Studie Gegenteiliges geschrieben steht. In diesen Fällen wird nach sorgfältiger Abwägung die Bewertung „teilweise erfüllt“ oder „nicht erfüllt“ vergeben. Einen Überblick über die CONSORT-Auswertungen geben Tabelle 8 im Anhang sowie zu ausgewählten CONSORT-Kriterien die Diagramme 2 und 3 auf den folgenden beiden Seiten. Wie in Gliederungspunkt 2.3 dieser Arbeit bereits beschrieben, bildet die Erfüllung der Forderungen des CONSORT-Statements die Grundlage für eine problemlose Bewertung einer Studie mithilfe des EBM-Schemas nach Sackett bzw. des objektivierten Bewertungssystems nach Hassler. Nur wenn in einer Publikation die Angaben vorhanden sind, auf die das Sackett'sche EBM-Schema zurückgreift, ist es möglich, zu einer auf Fakten gestützten Argumentation zu gelangen. Anderenfalls muss spekuliert werden, ob die einzelnen Punkte erfüllt wurden und im Zweifel gegen eine Studie entschieden werden. Besonders wichtig ist eine Erfüllung der CONSORT-Kriterien Nummer 8 und 9, die sich mit der Zufallszuteilung befassen, Nummer 13 und 16 für die Frage nach der Anzahl der berücksichtigten Patienten, Nummer 16 jedoch auch für die Bestimmung, ob es sich um eine Intention-to-treat-Analyse handelte, und Nummer 11, die sich mit der Verblindung befasst. Durch Erfüllung der CONSORT-Kriterien 3 und 4 wird eine Überprüfung der Gleichbehandlung der Patientengruppen möglich, und Nummer 15 dient der Frage, ob die Vergleichsgruppen einander ähnlich waren. Die anderen CONSORT-Forderungen machen zusätzliche Angaben einerseits zu den eben genannten klassischen EBM-Kriterien (EBM-Schritte 1-6), andererseits und besonders aber auch zu EBM-Schritt Nummer 7, der Frage nach zusätzlichen die Validität einschränkenden Aspekten. Bei einer Studie, die in mehreren Publikationen auf unterschiedliche Weise veröffentlicht wurde, ist eine Bewertung nach EBM-Kriterien jedoch nicht so streng vorzunehmen wie für die CONSORT-

Forderungen. Bei CONSORT ist es nämlich erforderlich, dass die notwendigen Angaben direkt in dem jeweils untersuchten Bericht gemacht werden, während hingegen für eine Bewertung nach Sackett/Hassler nur wichtig ist, dass die relevanten Informationen bei mehrfacher Publikation einer Studie in einer dieser Veröffentlichungen gefunden werden können. Dies wurde bei der in Punkt 3.1.2. dieser Arbeit beschriebenen Studienbewertung nach EBM-Kriterien berücksichtigt.

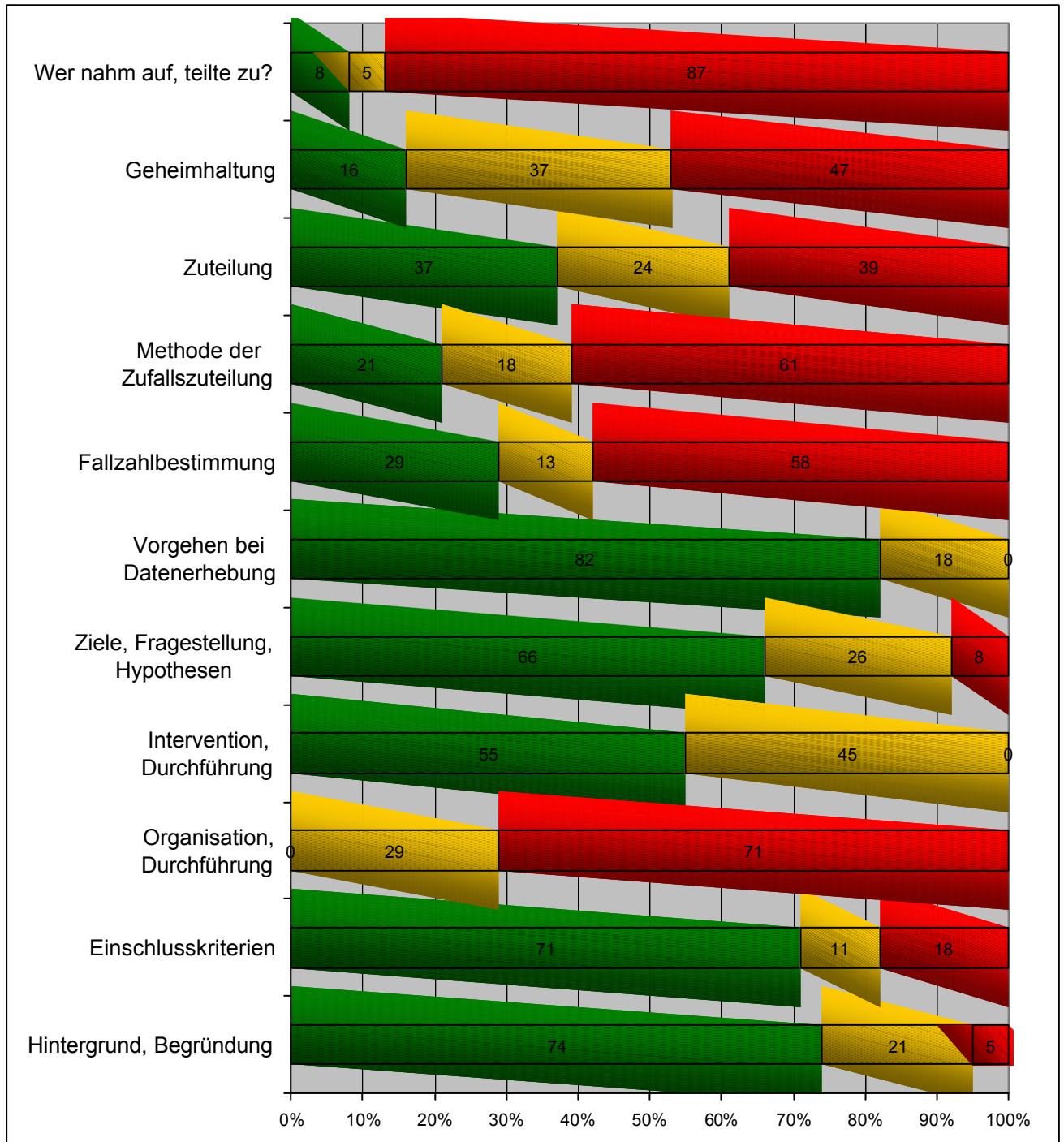


Diagramm 2: Erfüllung von CONSORT-Kriterien, Teil 1. Angaben in Prozent aller bewerteter Studien.

■ = Ausreichende Beschreibung.

■ = Zu knappe Beschreibung oder Verweis auf Schwesterstudie.

■ = Keine Beschreibung.

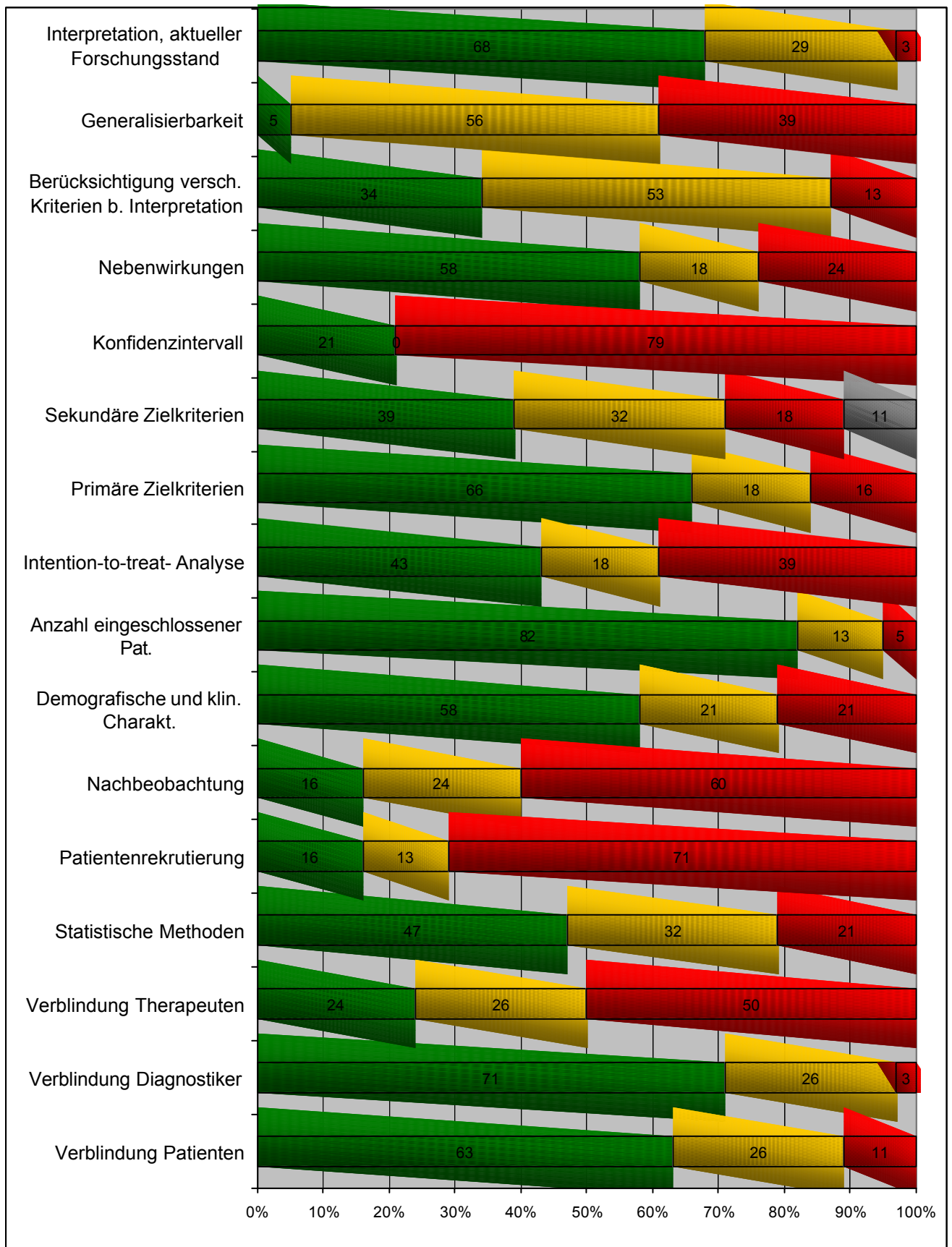


Diagramm 3: Erfüllung von CONSORT-Kriterien, Teil 2. Angaben in Prozent aller bewerteter Studien.

- = Ausreichende Beschreibung.
- = Zu knappe Beschreibung oder Verweis auf Schwesterstudie.
- = Keine Beschreibung.
- = Keine sekundären Zielkriterien verwendet.

3.1.1.2.1. Ergebnisse der Untersuchung von Titel, Zusammenfassung und Einleitung

Für Titel und Zusammenfassung fordert die CONSORT-Checkliste eine kurze Beschreibung der Patientenzuordnung zu den Behandlungsgruppen, also ob die Zuweisung randomisiert erfolgte. In 19 Studien (49%) war nur im Abstract von einer Randomisierung die Rede, sechs Studien (16%) erwähnten die Randomisierung in Zusammenfassung und Überschrift, eine Studie (3%) sprach von Randomisierung nur im Titel und nicht im Abstract, und in zwölf Studien (32%), also fast einem Drittel, war in keinem der beiden Bereiche eine Zufallszuteilung zu erkennen (siehe Diagramm 4).

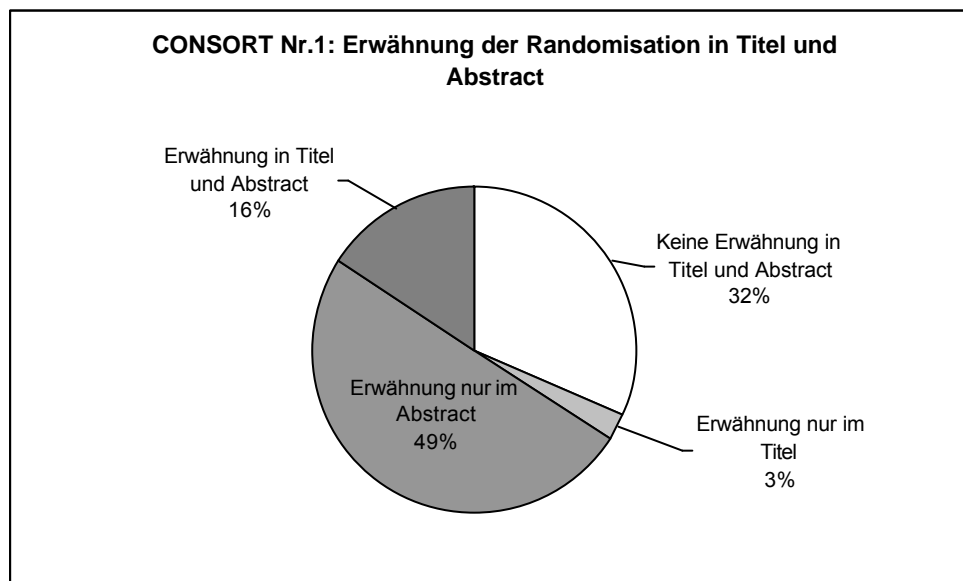


Diagramm 4: Erwähnung der Randomisation in Titel und Abstract.

Die Einleitung betreffend ist die Forderung des CONSORT-Statements, dass wissenschaftlicher Hintergrund und Begründung der Studie beschrieben sind. In fast $\frac{3}{4}$ der Fälle, das heißt 28 Studien oder 74%, konnte eine ausreichende Beschreibung festgestellt werden. Um nur ein Beispiel zu nennen sei hier auf Filippi 2000, S. 1714f. [7] verwiesen, wo zunächst der experimentelle Hintergrund von Cladribin erklärt wird, dann auf vorherige klinische Studien und deren Ergebnisse zu diesem Stoff verwiesen wird und schließlich die Gründe für eine erneute Studiendurchführung dargelegt werden. Entweder zu knapp erklärt beziehungsweise nur mit einem Verweis auf eine Schwesterstudie oder einen anderen Studienreport versehen waren immerhin acht Einleitungen (21%). Keine Begründung für die Studiendurchführung oder Beschreibung des wissenschaftlichen Hintergrunds fand sich in zwei Studien (5%). In Teksam 2000 [40] (S.885) wird zum Beispiel lediglich erklärt, dass MRI ein wichtiges Instrument zur Evaluation eines Behandlungserfolgs darstellt und abschließend in nur einem Satz erwähnt, dass das Vorhaben der Studie war, die Effizienz von intravenösem IgG zu bestimmen. Damit wird jedoch nur in zu

knapper Weise eine Begründung für die Studiendurchführung oder der wissenschaftliche Hintergrund genannt (siehe „Hintergrund, Begründung“ in Diagramm 2).

3.1.1.2.2. Ergebnisse der Untersuchung des Material-und Methoden-Teils

Die erste Forderung für den Material-und Methoden-Teil (CONSORT Nr.3) ist die nach der Beschreibung von Einschlusskriterien für Probanden, nach der Studienorganisation und nach dem Ort der Studiendurchführung. In der Untersuchung für die vorliegende Arbeit wurde gefunden, dass Ein-und Ausschlusskriterien in 27 Publikationen (71%) ausreichend beschrieben sind, wohingegen in vier RCTs (11%) die Beschreibung nur knapp oder nur in Schwesterberichten zu finden war. Sieben Veröffentlichungen (18%) machten keinerlei Angaben (siehe „Einschlusskriterien“ in Diagramm 2). Ein sehr gutes Beispiel für eine ausführliche Auflistung der Ein-und Ausschlusskriterien bietet Goodkin 2000 [10] (S. 1415). Auf eine direkte Zitation wird an dieser Stelle des Umfangs wegen verzichtet.

Deutlich war das Ergebnis zu der Berichterstattung über die Studienorganisation und den Ort der Patientendatenerhebung. Keine Studie berichtete darüber ausführlich genug. Elf Publikationen (29%) beschrieben Studienorganisation-und Ort zu knapp, und in 27 Veröffentlichungen (71%) waren keine ausreichenden Angaben enthalten (siehe „Organisation, Durchführung“ in Diagramm 2).

Bei der Beschreibung der geplanten Interventionen und deren Durchführung (CONSORT Nr.4) war das Ergebnis etwas günstiger, wobei auch hier einige Mängel zu finden waren: 21 Studien (55%) berichteten ausführlich genug, wobei hier bereits mit eingerechnet wurde, wenn ein Studienbericht knappe Angaben zur Studienmedikation machte und dabei auf einen Zweitstudienbericht verwies, in dem dann eine ausführliche Beschreibung erfolgte. 17 Publikationen (45%) machten zu knappe Angaben ohne Bericht in einer Schwesterpublikation oder enthielten die Daten nur in einer Zweitveröffentlichung. In keiner Studie fehlten die Angaben völlig (siehe „Intervention, Durchführung“ in Diagramm 2). Ein recht gutes Beispiel, wie in zwei Sätzen die wichtigsten Angaben gemacht werden können, bietet Noseworthy 1998 [27] (S.1343): „Patients were randomized to receive either enteric-coated sulfasalazine (Salazopyrin EN-tabs, 500 mg; Kabi Pharmacia AB, Uppsala, Sweden) or placebo. Patients took one tablet daily, and increased their dose by one tablet at weekly intervals to four tablets per day. ...“ Allerdings könnten durchaus noch mehr Angaben, wie zum Beispiel Art und Aussehen des Placebos oder die Uhrzeit der täglichen Medikamentengabe, vorhanden sein. Ein weiteres Beispiel zeigt Zhao 2000 [46] (S.201): „... patients ... were randomized into three treatment arms: placebo and 1.6 or 8.0 mIU IFN β -1b, self-administered subcutaneously on alternate days.“ Ansich wäre

diese Aussage zu knapp. So wären Informationen über die Art der subkutanen Injektion, ob die Patienten ein Injektions-Training erhielten oder Daten zum Hersteller der zu injizierenden Flüssigkeit, hilfreich. Da aber in „The IFNB Multiple Sclerosis Study Group“, *Neurology* 1993 [150] (S.656), diese Angaben vorhanden sind, wurden die Ausführungen hier als ausreichend angesehen.

Genauere Ziele, Fragestellung und Hypothesen (CONSORT Nr.5) wurden in 25 Studien (66%) beschrieben, zehn RCTs (26%) machten zu knappe Angaben und drei Publikationen (8%) waren in diesem Punkt mangelhaft (siehe „Ziele, Fragestellung, Hypothesen“ in Diagramm 2). In einer dieser drei Studien war die einzige Angabe zu diesem Punkt: „It is not known, ..., whether rIFN- β_{1a} treatment also modifies the pathological course of individual new lesions that develop during treatment. This study used serial MRI to investigate whether rIFN- β_{1a} treatment reduces the duration of Gd enhancement and affects the short-term outcome of new enhancing lesions.“ Die wichtigsten Angaben, das sind die Ziele einer Studie, lassen sich aus diesem Satz zwar einigermaßen entnehmen, aber Hypothesen oder nähere Ausführungen dazu sind nicht enthalten.

Die nächste Forderung, CONSORT-Nr.6, ist die Beschreibung klar definierter primärer und sekundärer Zielkriterien und aller zur Optimierung der Ergebnisqualität verwendeten Methoden. Für Ergebnisse der Beantwortung von Fragen oder Forderungen aus CONSORT-Statement oder Sackett-Schema muss wieder darauf hingewiesen werden, dass diese zum Teil, wie in diesem Punkt der CONSORT-Checkliste, vom Bewerter abhängig sind. Unterschiedliche Beurteiler setzen verschiedene Maßstäbe. Deshalb ist es wichtig, zu erläutern, wie die Resultate der CONSORT-Forderung 6.1 für diese Arbeit zustande kamen: Die Forderung gilt als voll erfüllt, wenn primäres und sekundäres Zielkriterium eindeutig gekennzeichnet waren, wie zum Beispiel in Jacobs 2000 [13] (S.899): „The primary prespecified end point was the development of clinically definite multiple sclerosis. ... Findings on MRI of the brain served as a secondary prespecified end point. ...“. Wenn eine eindeutige Zuordnung als Primär- oder Sekundär-Outcomekriterium nur in einem Nebenbericht erwähnt ist oder eine eindeutige Beschreibung zwar fehlt, aber trotzdem aus den Angaben erkennbar ist, welche Messungen wahrscheinlich als primäre und sekundäre Zielkriterien dienten, wurde dies mit „zum Teil erfüllt“ bewertet. Die Beurteilung wird verkompliziert durch die Tatsache, dass manche Studien Primärkriterien definieren, aber Sekundärkriterien nicht eindeutig kennzeichnen. Dies wurde als zusätzliche Bewertungsstufe eingeführt. Als nicht erfüllt wurde die Forderung angesehen, wenn unklar war, welche Beurteilungen besonders wichtig und welche zweitrangig waren. Es reichte nicht aus, nur die Vorgehensweise bei den für verschiedene Zielkriterien notwendigen Untersu-

chungen zu beschreiben. Eine Beschreibung der Vorgehensweise bei der Evaluation verschiedener Zielkriterien ist ein zusätzlicher von CONSORT Nr.6 geforderter Punkt, der unabhängig von der klaren Definition primärer und sekundärer Zielkriterien ist.

Die Überprüfung der Zielkriterienbeschreibung erbrachte folgendes Ergebnis: In 15 Studien (40%) waren Primär- und Sekundärkriterien eindeutig gekennzeichnet. Ein sehr gutes Beispiel dafür bietet Millefiorini 1997 [19] (S.154), wo die Endpunkte eindeutig aufgelistet sind:

„Efficacy end-points:

Treatment efficacy was assessed by comparing the following variables between the placebo and MTX group.

Primary end-point

1. Proportion of patients with confirmed progression as measure by an increase of at least one point on the EDSS scale.

Secondary end-points

1. Annual mean number of exacerbations and proportion of exacerbation-free patients.

2. Change in mean EDSS from baseline to end-point.

3. Mean number of new or enlarged lesions on T2-weighted MRI performed at baseline and 12 and 24 months after the beginning of the study.”

Fünf Publikationen (13%) benannten zwar das Primärkriterium, ob die weiteren Messungen als primär oder sekundär zu gelten hatten, blieb jedoch unklar. In weiteren fünf RCTs (13%) war eine Kennzeichnung nur in einer Schwesterstudie zu finden, oder es konnte aus den Angaben der bewerteten Studie die Priorität der Kriterien zumindest teilweise erschlossen werden. 13 Publikationen (34%) kennzeichneten die Rangfolge ihrer Zielkriterien nicht (siehe Diagramm 5).

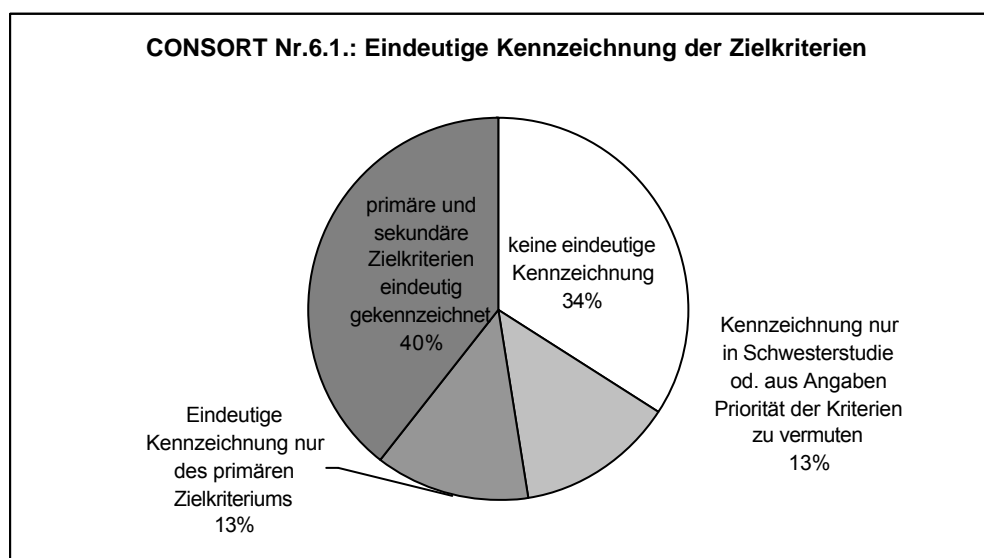


Diagramm 5: Eindeutige Kennzeichnung der Zielkriterien.

Trotz zum Teil fehlender Beschreibung der Wertigkeit von Zielkriterien machten 31 Studien (82%) genügend Angaben, um das jeweilige Vorgehen bei der Datenerhebung nachvollziehen zu können, und waren unabhängig davon, ob die Zielkriterien eine Kennzeichnung als primär oder sekundär hatten. Ein Beispiel für eine Beschreibung der Vorgehensweise zeigt Filippi 2000 [7] (S.1715f.), wo erklärt wird, wie Gehirnvolumina gemessen wurden. In sieben Publikationen (18%) waren die Angaben zu der Vorgehensweise bei der Datenerhebung knapp, keine ließ den Leser vollkommen im Unklaren. Eine eindeutig zu knappe Ausführung dieser Informationen bietet Teksam 2000 [40] (S.885f.). Dort finden sich nur sehr kurze und nur die allernötigsten Angaben, jedoch keine Zusatzinformationen (siehe „Vorgehen bei Datenerhebung“ in Diagramm 2).

Im Zusammenhang mit CONSORT-Forderung Nummer 6 wurde für die vorliegende Arbeit auch untersucht, ob die für die Bewertung der MRT-Bilder zuständigen Ärzte ein Training für die Auswertung erhielten. Leider war in 31 Studien (82%) kein zusätzliches auf die jeweils durchgeführten MR-Untersuchungen zugeschnittenes Training erfolgt oder es fehlt die Angabe. Lediglich in fünf Publikationen (13%) wurde beschrieben, wie die Auswerter auf ihre Aufgabe vorbereitet worden waren. Zwei RCTs (5%) erhielten eine Zwischenwertung: In einer Studie war die Angabe zu knapp und nur beschrieben, dass überhaupt ein Training durchgeführt worden war (Rice 2000, S.1147 [32]) und in einer weiteren RCT war ein Training wegen objektiver Bewertungsmethoden offensichtlich überflüssig (Molyneux 2000, S.2257 [22]) (siehe Diagramm 6).

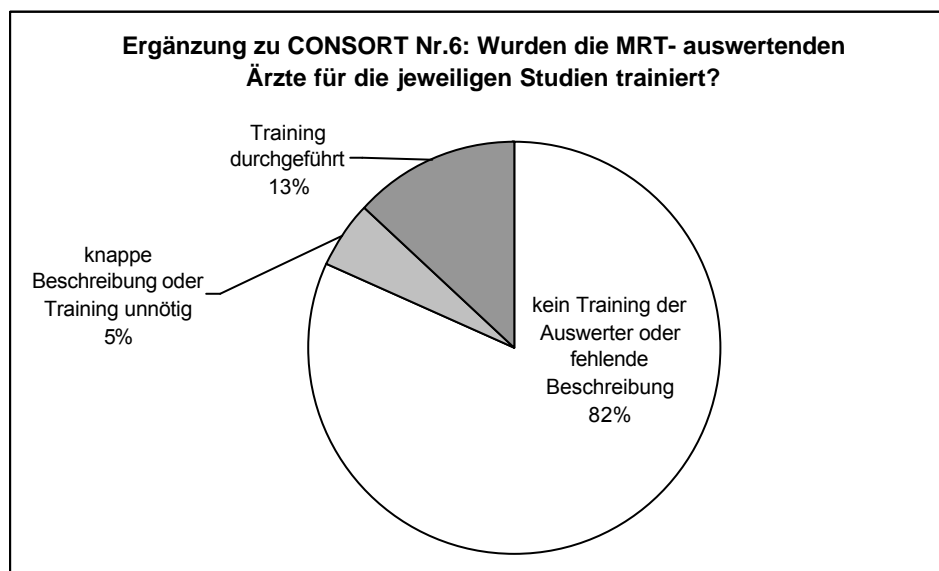


Diagramm 6: Training der auswertenden Ärzte durchgeführt und beschrieben?

Erwähnenswert bei CONSORT Nummer 6 ist schließlich noch, dass es wichtig ist, in einer Studie bereits etablierte Auswertungsmethoden zu verwenden. Dabei ist als positiv zu vermerken, dass alle eingeschlossenen Studien mit klinischen Outcomes den EDSS-Score zumindest als eines von mehreren Kriterien verwendeten, womit eine Vergleichbarkeit innerhalb der RCTs erleichtert wird.

Der erste Teil in Forderung Nummer sieben der CONSORT-Checkliste ist die genaue Beschreibung, wie die für eine Studie benötigten Fallzahlen bestimmt wurden. 22 Veröffentlichungen (58%) machten keine Angaben dazu, fünf Studienberichte (13%) enthielten immerhin knappe Daten, und in elf Publikationen (29%) gab es ausreichende Angaben, um die Fallzahlberechnung nachvollziehen zu können (siehe „Fallzahlbestimmung“ in Diagramm 2), wie zum Beispiel in Myhr 1999 [25] (S.1050), wo sich eine sehr gute Erklärung der Fallzahlberechnung findet: „The number of patients included was based on power calculations assuming that 15% of RRMS patients have no active lesions on MRI during 6 months without treatment. For a one-sided p value of 0.05, 28 patients in each arm are needed to reach a power of 80% in detecting a threefold increase of this proportion among treated patients. ...“

Der zweite Teil der siebten Forderung verlangt eine Beschreibung von Zwischenanalysen und Abbruchkriterien. Neun Studien (24%) gaben diese ausreichend an, zehn Publikationen (26%) machten nur knappe Angaben sowohl zu Zwischenanalysen als auch zu Abbruchkriterien oder gaben nur eines der beiden an, und zehn Veröffentlichungen (26%) beschrieben weder Zwischenanalysen, obwohl durchgeführt, noch Abbruchkriterien in adäquater Weise. In neun RCTs (24%) gab es keine Zwischenanalysen und damit auch keine Angaben dazu (siehe Diagramm 7). Von diesen wurden nur in einer (11% von neun Publikationen) Abbruchkriterien genau angegeben (Myhr 1999, S.1051 [25]).

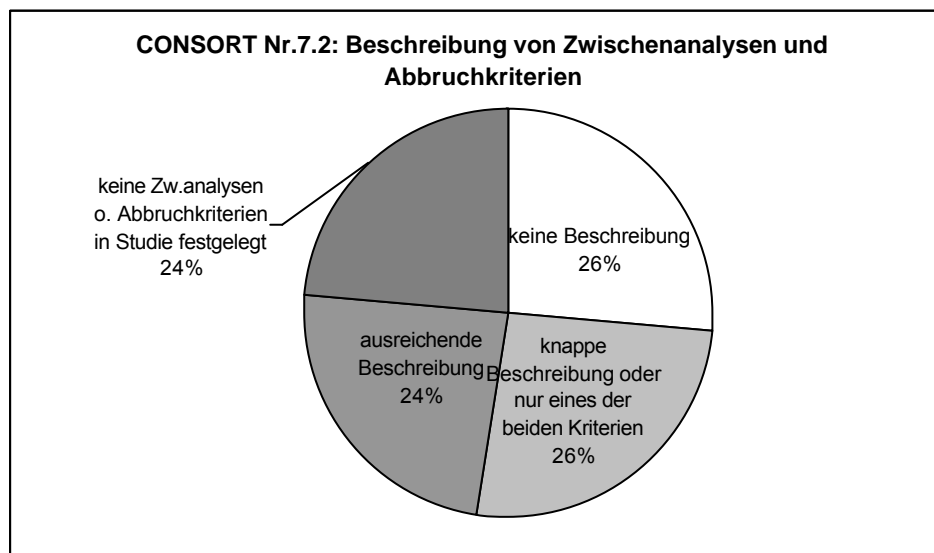


Diagramm 7: Beschreibung von Zwischenanalysen und Abbruchkriterien.

Nummer acht auf der CONSORT-Liste verlangt nach einer genauen Beschreibung der Randomisationsmethode. Diese war nur in acht Studien (21%) gegeben. Ein sehr gutes Beispiel für exakte Angaben zur Art der Zufallszuteilung bietet „The Lenercept Multiple Sclerosis Study Group“ 1999 (S.458 [41]): „A randomization list with treatment blocks (four patients per block) was computer generated by Hoffmann-La Roche (Basel, Switzerland) for each investigation site. During the conduct of the study, the randomization list was

available only to the Safety Review Board ... members ...“. Sieben Publikationen (18%) verwiesen mit einer Fußnote auf die Beschreibung in einem anderen Studienbericht und in 23 RCTs (61%) fehlte eine Angabe über die Methode der Zufallszuteilung (siehe „Methode der Zufallszuteilung“ in Diagramm 2).

In der folgenden Forderung (Abschnitt 9) wird nach der Durchführung der Patientenzuteilung gefragt. Es soll erläutert sein, ob in einer Studie zum Beispiel nummerierte Blätter verwendet wurden oder ob eine zentrale Randomisation stattfand. Außerdem muss, so der zweite Teil der neunten Forderung, angegeben sein, ob die Geheimhaltung bis zur Zuteilung gewährleistet war.

Ausreichende Angaben über die Durchführung der Zuteilung waren in 14 Publikationen (37%) vorhanden, neun Veröffentlichungen (24%) verwiesen dazu auf einen anderen Studienbericht, in dem die Beschreibung zu finden war, und 15 Publikationen (39%) machten keine adäquaten Aussagen (siehe „Zuteilung“ in Diagramm 2).

Leider waren über die Geheimhaltung der Zufallszuteilung nur wenige Veröffentlichungen aussagekräftig – lediglich sechs (16%). 14 Studienberichte (37%) machten zumindest knappe Angaben oder verwiesen auf einen Zweitstudienbericht bzw. auf eine Schwesterstudie mit genauerer Beschreibung, und mit 18 Publikationen oder 47% war knapp die Hälfte in der Geheimhaltung insuffizient oder berichtete zumindest nicht entsprechend darüber (siehe „Geheimhaltung“ in Diagramm 2).

Goodkin beschreibt die Durchführung der Zuteilung folgendermaßen und bietet damit ein gutes Beispiel für die Erfüllung beider Teile von Forderung 9: "Randomization was performed centrally, with the allocation sequence computer generated and concealed in sequentially numbered, opaque envelopes until the patient was randomly assigned. Only the pharmacist at each site and the two members of the independent Data and Safety Monitoring Board (DSMB) were aware of treatment assignment." (Goodkin 2000, S.1415 [10]). Ein weiteres Exempel für eine gute Beschreibung der Durchführung der Zuteilung zeigt "The Once Weekly Interferon for MS Study Group" 1999 (S.680 [42]): „Treatment assignment was determined by a computer-generated randomization list produced by the Corporate Biometrics Department of Ares-Serono. The randomization code for each patient was delivered to the investigator in sealed envelopes to be opened only in emergency situations... . Envelopes were returned at the end of treatment. Patients were randomized in a 1:1:1 ratio, stratified by study center, to receive IFN β -1a ...“.

Eine weitere Forderung, die Zuteilung der Patienten zu den Behandlungsgruppen betreffend, ist die Beschreibung, wer die Randomisationssequenz herstellte, wer die Probanden in die Studie aufnahm und wer die Patienten ihrer Gruppe zuwies. Dies ist Nummer 10 der

CONSORT-Liste. Leider konnte eine ausreichende Beschreibung nur in drei Studien (8%) gefunden werden. Zwei Publikationen (5%) machten zumindest knappe Angaben, und der Großteil von 33 Veröffentlichungen (87%) enthielt keine entsprechenden Ausführungen (siehe „Wer nahm auf, teilte zu?“ in Diagramm 2). Ein Beispiel für einen Teil dieser Forderung bietet „The Once Weekly Interferon for MS Study Group“ 1999 (S.680 [42]), wie unter CONSORT Nummer 9 zitiert. Ein anderes Exempel für einen wichtigen Teil der Forderung zeigt Cook 1997 (S.174 [3]): „... the patient was randomized to receive sham TLI or TLI by the radiation oncologist opening a sealed assignment envelope previously prepared by the biostatistician.“ Hier wird also deutlich gemacht, wer den Umschlag öffnete und damit verantwortlich war. Diese Studie war die einzige, in der der direkt Zuständige genannt wurde.

Ein besonders wichtiger Punkt ist Nummer 11 der CONSORT-Checkliste – die Verblindung. Es ist notwendig, dass sowohl die an einer Studie teilnehmenden Patienten (Nr.11a) und Diagnostiker (Nr.11c), aber auch die Therapeuten (Nr.11b) nicht über die Behandlung informiert sind. CONSORT fragt dabei, ob dies ausreichend berichtet wird, wobei natürlich auch ein eindeutiger Bericht über eine fehlende Verblindung als gut beschrieben angesehen wird, während dies im EBM-Schema negativ bewertet wird.

Kriterium 11a wurde in 27 Publikationen (63%) gut nachvollziehbar erfüllt, und in acht Veröffentlichungen (26%) war der Bericht über eine Verblindung der Patienten entweder in der bewerteten Publikation knapp oder nur in einem anderen Report über die selbe Studie zu finden. Wie die Verblindung gesichert wurde, ist zum Beispiel in Oliveri 1998 [28] gut beschrieben: „To ensure the integrity of the blinding, all IVIG and albumin solutions were prepared in the hospital pharmacy and the container and infusion tubes were covered with red plastic wrappers.“ (Oliveri 1998, S.1274 [28]) Hohe Transparenz im Bereich der Patientenverblindung bietet auch Beck 1993 [62]: "Whereas the patients in the oral-prednisone and placebo groups were not informed of their treatment assignments, those in the intravenous-methylprednisone group were aware of their assignments." (Beck 1993, S.1765 [62]). Diese Studie konnte im EBM-Schema nicht die volle Punktzahl im Bereich der Verblindung erhalten, weil ja ein Teil der Patienten nicht verblindet war. Die Forderung 11a) des CONSORT-Statement wurde jedoch erfüllt, da berichtet wird, welche Patienten verblindet waren und bei welchen keine Verblindung gegeben war. Vier RCTs (11%) schließlich machten keine expliziten Erklärungen zu einer Patientenverblindung (siehe „Verblindung Patienten“ in Diagramm 3).

Die Verblindung der diagnostisch tätigen Ärzte war sogar in einer noch höheren Zahl von Studien gut beschrieben: 27 Publikationen (71%) enthielten gute oder sehr gute Erklärungen zur Verblindung der Diagnostiker, in zehn Studien (26%) sind die Informationen ent-

weder nicht ausführlich genug oder nur in einem Schwesterbericht zu finden und in einer Veröffentlichung (3%) fehlen Angaben zu einer Diagnostikerverblindung (siehe „Verblindung Diagnostiker“ in Diagramm 3).

Leider sind die Zahlen für Therapeuten, auch wenn in diesem Bereich nicht ganz so wichtig wie für Patienten und Diagnostiker, weniger positiv. Nur in 9 Studien (24%) ist gut nachzuvollziehen, ob die therapeutisch tätigen Ärzte verblindet waren oder nicht. Zehn RCTs (26%) verweisen auf einen parallelen Studienbericht bzw. sind in der bewerteten Publikation zu knapp gehalten, und in 19 Veröffentlichungen (50%) fehlt die Angabe (siehe „Verblindung Therapeuten“ in Diagramm 3) Einen vergleichenden Überblick der Verblindung bietet Diagramm 8.

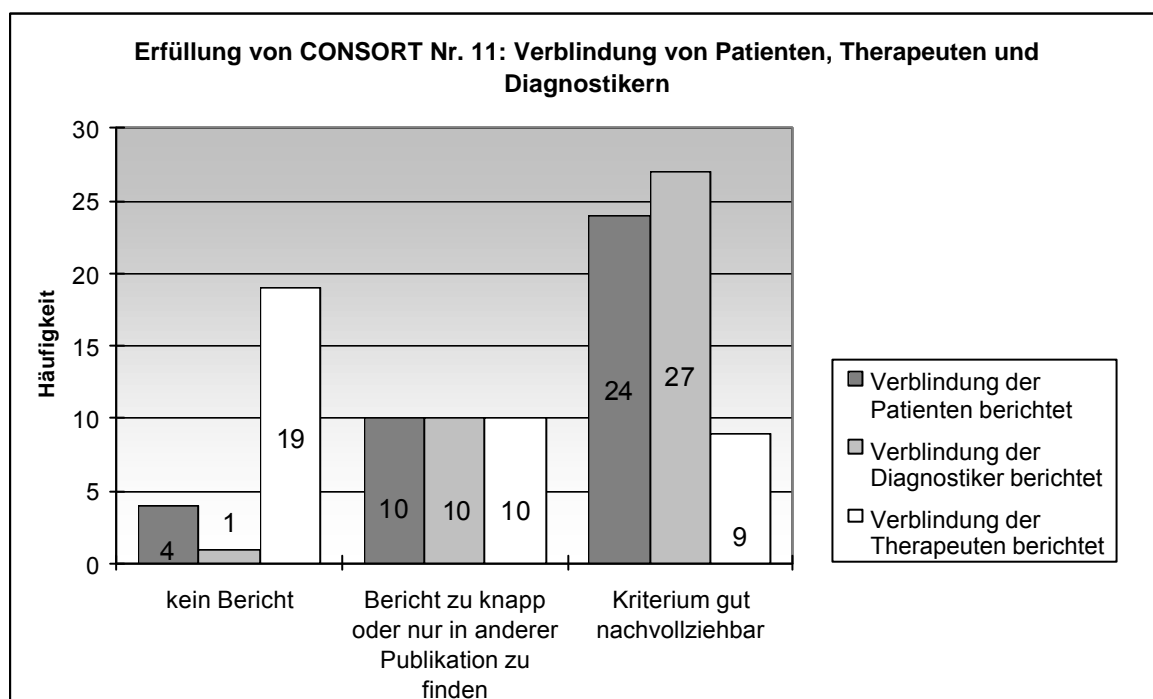


Diagramm 8: Erfüllung von CONSORT-Forderung Nummer 11: Beschreibung der Verblindung von Patienten und Ärzten. Die Zahlen in den Balken geben die Anzahl der Studien wieder.

Ein sehr gutes Beispiel für einen Bericht über die Verblindung sowohl von Ärzten als auch Patienten bietet Noseworthy 1998 [27] (S.1348): „A ‘forced-choice’ approach was used to assess the adequacy of the blinding. One evaluating physician and six treating physicians believed that they had become unblinded, but four of the six unblinded treating physicians guessed incorrectly, indicating that they truly were blinded. The physicians were well blinded in that approximately 50% of the sulfasalazine-treated patients were thought to be on placebo (49% evaluating physicians, 52% treating physicians) and a significant percentage of placebo-treated patients were misclassified (39% evaluating physicians, 48% treating physicians). A total of 55% of the sulfasalazine-treated patients guessed their treatment assignment correctly and 50% of placebo-treated patients guessed correctly. ...” Rice 2000 [32] schreibt auf S.1147: „The treating physician remained blinded but was

provided with any abnormal CBC results required for proper medical management.“ Für diese Angaben konnte der Publikation nur ein Punkt in der CONSORT-Auswertung zugesprochen werden, da zwar die Angaben nicht völlig fehlen, jedoch auch nicht mit Sicherheit auf eine komplette oder fehlende Verblindung geschlossen werden kann. Eine Angabe wie „The treating physician, not blinded to study treatment ...“ (Millefiorini 1997, S.154 [19]) wird in der Bewertung nach CONSORT wiederum mit zwei Punkten bewertet, da die Beschreibung über eine Therapeutenverblindung vorhanden ist. Nach EBM erhält diese Studie in diesem Bereich natürlich null Punkte, da eine Verblindung nicht gegeben war.

Der Bericht über die angewandten statistischen Methoden ist die letzte Forderung der CONSORT-Gruppe für den Methodenteil. Eine Beschreibung war in 18 Veröffentlichungen (47%) zu finden, in zwölf Studien (32%) waren die Angaben nicht ausführlich genug oder es war auf eine andere Publikation derselben klinischen Studie verwiesen, und in acht RCTs (21%) gab es keinen Bericht über statistische Methoden (siehe „statistische Methoden“ in Diagramm 3). Ein Beispiel, wie eine Beschreibung aussehen könnte, zeigt Rudick 1999 (S.1700 [34]). Dort werden die verwendeten Methoden genannt und dabei erklärt, wofür diese eingesetzt wurden: „Student’s *t*-tests were used to compare baseline and percent changes in BPF between placebo and IFN β -1a treatment groups, and between MS patients and healthy controls. Effect of treatment was evaluated using analysis of covariance adjusting for baseline covariates. ...“. Ganz anders zum Beispiel die Darstellung in Jacobs 2000 [12], wo überhaupt keine Angaben gemacht werden, oder Coles, wo die Beschreibung der statistischen Methoden sich auf „Spearman’s rank correlation test was used to test the data“ (Coles 1999, S.298 [2]) beschränkt.

3.1.1.2.3. Ergebnisse der Untersuchung des Ergebnisteils

Für den Ergebnisteil enthält die CONSORT-Kontrollliste sieben Forderungen (siehe Abbildung 3 auf Seite 17). Die erste ist eine Darstellung des Patientenflusses durch eine Studie, am besten mithilfe eines Flussdiagramms. Daher wurde für die vorliegende Arbeit untersucht, in wie vielen Publikationen ein Flussdiagramm enthalten ist. Es konnte gefunden werden in 29 Studien (76%), neun Veröffentlichungen (24%) enthielten keines (siehe Diagramm 9). Beispiele für sehr gut aufgebaute Flussdiagramme finden sich unter anderem in The Once Weekly Interferon for MS Study Group 1999 [42] auf Seite 682 (siehe Abbildung 11) und in The Lenercept Multiple Sclerosis Study Group 1999 [41] auf Seite 460 (siehe Abbildung 12).

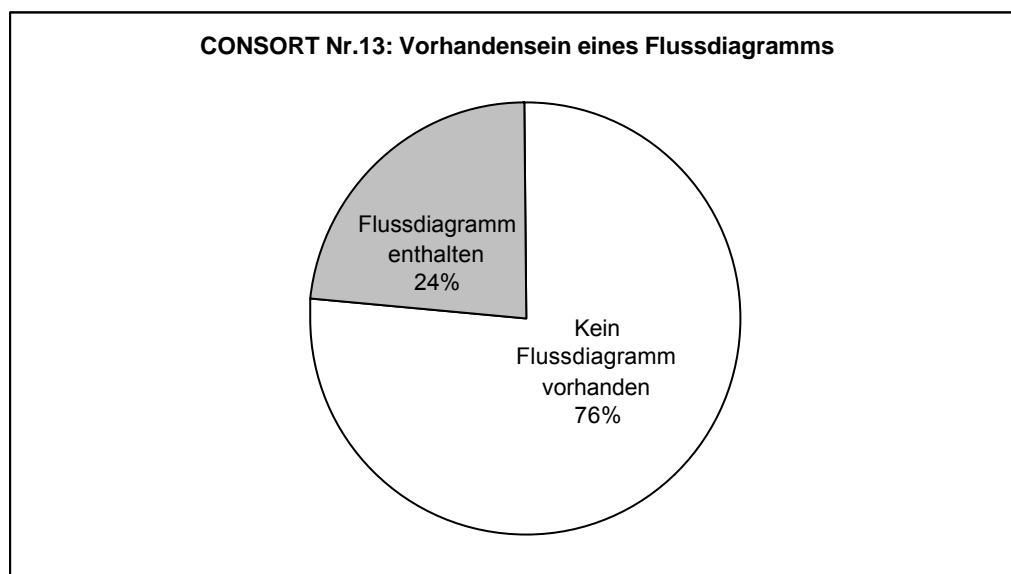


Diagramm 9: Vorhandensein eines Flussdiagramms.

Der nächste Punkt auf der CONSORT-Liste fragt nach näheren Angaben zu der Zeit der Patientenaufnahme und der Nachbeobachtung. Wichtig ist dabei neben der Rekrutierungsrate vor allem die Auflistung der genauen Daten, also in welchem Zeitraum die Patienten rekrutiert wurden und die Nachbeobachtungsphase durchlaufen wurde (Altman 2001, S.679f. [48]). Für die vorliegende Arbeit wurden beide Forderungen getrennt voneinander betrachtet. Folgende Daten zur Rekrutierung konnten gefunden werden: Nur sechs Studien (16%) enthielten ausreichende Angaben zu der Rekrutierung und fünf Publikationen (13%) zumindest teilweise Daten bzw. Ausführungen in Nebenveröffentlichungen. Mit 27 RCTs (71%) enthielten fast $\frac{3}{4}$ aller Studien keine oder fast keine Daten zu den Zeiträumen und der Rate der Patientenrekrutierungen (siehe „Patientenrekrutierung“ in Diagramm 3). Nur gering günstiger sind die Auswertungsergebnisse zu den Angaben der Nachbeobachtungsphase. Sechs Studien (16%) enthielten hier ausreichende Informationen und neun Publikationen (24%) wenigstens knappe Daten. Mit 23 Studien 60% aller untersuchten RCTs vermittelten keine oder zu wenige Angaben zu der Durchführung der Nachbeobachtungsphase (siehe „Nachbeobachtung“ in Diagramm 3).

Gut beschrieben wurde der Zeitraum der Rekrutierung zum Beispiel in Li 1999 [18] (S.198): „A total of 560 patients with clinically definite or laboratory-supported RRMS of at least 1 year’s duration were enrolled between May 1994 and February 1995.“ Jacobs 2000 [13] (S.899), wählt folgende Formulierung: „The study was conducted at 50 clinical centers in the United States and Canada from April 1996 until March 2000.“

Für die Angaben zur Nachbeobachtungsphase schreibt Noseworthy 1998 [27] (S.1346): „Patients were treated and followed on medication until the final randomized patient had been in the study for 36 months. The mean follow-up was 3.7 years (range, 1.0 to 4.6 years). All patients completed 1 year of follow-up; 97% remained in the study for at least 2

years, and 94% for at least 3 years ...". Nachdem außerdem die Zeitpunkte der Randomisation genau angegeben waren, lässt sich daraus genau auf die Nachbeobachtungsphase schließen. Noch besser wäre es allerdings, wenn auch das Datum von Anfang und Ende der Follow-up-Zeit exakt genannt wäre. Ebenfalls nicht perfekt, aber ausreichend beschrieben und damit als „Kriterium erfüllt“ bewertet, wird in Myhr 1999 [25] (S.1050) die Dauer der Nachbeobachtungsphase genannt: „The patients were randomized to receive placebo ... or ... IFNa2a ... three times weekly for 6 months, followed by a 6-month follow-up.“ Auch wenn der Leser damit den zeitlichen Kontext der Studie nicht exakt abschätzen kann (Altmann 2001, S.680 [48]), da die Daten der Studiendurchführung nicht genannt werden, wird er doch in die Lage versetzt, den Studienablauf nachvollziehen zu können.

Nummer 15 der CONSORT-Checkliste fordert, dass sämtliche demografischen und klinischen Charakteristika aller Gruppen angegeben werden. Es ist nicht ausreichend, nur Ein- und Ausschlusskriterien anzugeben, sondern es muss auch eine Gegenüberstellung der Patientendaten in den Vergleichsgruppen zu Studienbeginn möglich sein. Am besten lässt sich dies in einer Tabelle darstellen (Altmann 2001, S.680f. [48]). Bei den in die Auswertung eingeschlossenen Studien war in 22 Publikationen (58%) der Vergleich der Patientendaten zu Studienbeginn uneingeschränkt möglich und waren in weiteren acht Studien (21%) zumindest in der Mutter- oder Schwesterstudie die nötigen Informationen enthalten (sechs Studien) bzw. die Angaben in der bewerteten Studie relativ knapp (zwei Studien). Keine Daten zu den eingeschlossenen Patienten gab es immerhin in acht Studien (21%) (siehe „Demografische und klin. Charakt.“ in Diagramm 3).

Jacobs 2000 [13] (S.901) bietet zum Beispiel eine sehr gute Tabelle, in der zusätzlich zu Alter, Geschlecht und Dauer der Erkrankung unter anderem nach Rasse und MR-Parametern unterschieden wird (siehe Abbildung 6). Ein weiteres Exempel für die Darstellung der relevanten Patientendaten zeigt The Lenercept Multiple Sclerosis Study Group 1999 [41] (S.460). Die dort gezeigte Tabelle kann zwar nicht die Ausführlichkeit der Darstellung von Jacobs erreichen, bietet jedoch einen ausreichenden Überblick über die Patientendaten (siehe Abbildung 13). Weniger gut dargestellt sind die Informationen zum Beispiel in Jacobs 1995 [93] (S.126), einem Vorabbericht unter anderem zu Rudick 1999 [34]. In der Table 3 sind zwar die Patienten unter anderem nach Ort der Datenerhebung, Geschlecht und Alter aufgelistet, allerdings lässt diese Tabelle keinen Vergleich der Behandlungsgruppen zu, da keine Vergleichsgruppen gegenübergestellt werden, sondern nur das Gesamtkollektiv dargestellt wird (siehe Abbildung 14). Diese Tabelle wurde als Beispiel für die vorliegende Arbeit herangezogen, da sie als Erläuterung dafür dienen kann, wie eine Darstellung nicht erfolgen sollte und welche Probleme auftreten können. Da es sich bei Jacobs 1995 jedoch nur um einen Vorabbericht für eine unter anderem in

Rudick 1999 [34] publizierte Studie handelt, fällt die fehlende Gegenüberstellung der Behandlungsgruppen in diesem Fall nicht ins Gewicht. Es wird lediglich die Aussagekraft einer derartigen Tabelle stark vermindert, weswegen selbst in Vorberichten eine Gegenüberstellung der Therapiegruppen erfolgen sollte. Da die Patientendaten jedoch in Rudick 1999 [34] (S.1700, Table 1) gut aufgeschlüsselt sind, stellt die fehlende Gegenüberstellung der Therapiegruppen hier kein Problem dar.

Um ein Beispiel für eine nicht ausreichende Beschreibung zu nennen, sei an dieser Stelle Tan 2000 [39], S.99 zitiert. In dieser Publikation fehlt eine Tabelle 1, und auch im Text sind die Angaben zu den Patienten bei weitem nicht ausreichend: „... 501 patients were to be enrolled in the 147 Line ... and 350 in the 146 Line trial Key inclusion criteria were clinically definite, or laboratory supported definite MS ... with ages between 18 and 60 (inclusive) years old. Patients were randomised by centre and stratified for baseline disability.“ (Tan 2000, S.99 [39]). Der Leser kann also in diesem Fall weder die jeweiligen beiden Behandlungsgruppen vergleichen noch überhaupt eine Aussage zu den eingeschlossenen Patienten machen, abgesehen von deren Altersumfang. Aufgetretene Wirkungen oder Nebenwirkungen des untersuchten Medikaments lassen sich also, da der Erfolg der Stratifizierung und Randomisierung unklar und damit nicht sicher ist, ob die Therapiegruppen gleiche Merkmale hatten, nicht eindeutig dem Arzneimittel oder möglichen Unterschieden im Patientenkollektiv zuordnen.

Als nächstes werden die bewerteten Studien auf die Angabe der Anzahl in die Analyse eingeschlossener Patienten untersucht und aufgezeigt, ob die Auswertung einer „Intention-to-treat“-Analyse folgte. Die Anzahl am Ende noch eingeschlossener Patienten war in 31 Publikationen (82%) und weiteren fünf RCTs (13%) zumindest knapp oder in einem parallelen Studienbericht enthalten (siehe „Anzahl eingeschlossener Pat.“ in Diagramm 3). Lediglich zwei Studien (5%) machten keine Angaben zu der Anzahl der am Ende ausgewerteten Patienten, aber in 15 Publikationen (39%) war nicht zu erkennen, ob die Analyse einer „Intention-to-treat“ folgte. Sieben Veröffentlichungen (18%) enthielten dazu immerhin Beschreibungen in Schwesterberichten und 16 Studien (43%) gaben eine „Intention-to-treat“-Analyse direkt an (siehe „Intention-to-treat- Analyse“ in Diagramm 3).

Ein gutes Beispiel für die Angabe der am Ende noch eingeschlossenen Patienten bietet Edan 1997 [4] (S.114): „During treatment, five patients dropped out at months 3,4, or 5, because of pronounced deterioration. All were in the steroid only group, and had highly active disease both on clinical and MRI criteria.“ Der Leser erfährt also die genaue Zahl der nicht mehr berücksichtigten und damit auch der ausgewerteten Patienten. Was bei dieser Publikation aber besonders hervorzuheben ist, ist Table 2 (Edan 1997, S.114 [4]) (siehe Abbildung 15). Dort werden die ausgeschlossenen Studienteilnehmer nochmals

einzelnen aufgezehlt nach Zeitpunkt des Ausscheidens, EDSS-Scores und Läsionen im MRT. Damit wird dem kritischen Leser ermöglicht, seine eigenen Schlussfolgerungen zu ziehen und zu bewerten, ob die Dropouts unter Umständen das Ergebnis verfälschen könnten. Unter anderem ebenfalls positiv zu verzeichnen ist die Beschreibung in Patti 1999 [31] (S.285): „No patients dropped out of the study. Blinding was lost for 8 patients on nIFN- β (3 in the RR group and 5 in the SP group) within the first month because of fever, none at later times. However, all 8 in whom blinding was lost were retained for the statistical analysis since the examining neurologist and the radiologist remained blinded even though the treating neurologist and the patients were not.“ Der Leser erhält damit die Information, dass trotz fehlender Verblindung die Patienten weiterhin in die Analyse eingeschlossen wurden und kann damit Schlussfolgerungen auf die Validität der Studie ziehen, in diesem Fall bezogen auf Schritt 4 (Verblindung) des Sackett-Fragebogens.

Zu der Frage nach einer ITT-Analyse sei an dieser Stelle Cook 1997 [3] (S.175) genannt: „In the event of withdrawal or drop out for any reason, patients were considered in the statistical analysis on an intent-to-treat basis on the basis of EDSS score at the point of withdrawal. There was no distinction in the statistical analysis among the various kinds of withdrawal.“ Hier wird also eindeutig beschrieben, dass und auch in welcher Form die ITT-Analyse stattfand. Ebenfalls als voll erfüllt wurde im folgenden Beispiel das Kriterium „Angabe, ob ITT-Analyse“ betrachtet. Tan 2000 [39] schreibt auf Seite 100: „... the patients who had a target dosage of 5mg daily have been analysed according to the true treatment dose in the first 3 months and not as an `intention to treat study´.“ In dieser Studie wurde also keine ITT-Analyse durchgeführt und in der EBM-Bewertung nach Sackett müsste Schritt 3 folgerichtig als nicht erfüllt betrachtet werden, das CONSORT-Kriterium Nummer 16.2 kann jedoch zustimmend beantwortet werden, da eine eindeutige Angabe über das *Fehlen* einer ITT-Analyse vorhanden ist. Eine mangelnde Beschreibung über die Auswertung nach ITT zeigt exemplarisch Oliveri 1998 [28] (S.1835): „Two patients ... failed to complete the study because a new exacerbation occurred during the follow-up Both patients started a new treatment course and were excluded from the analysis.“ Hier wird also deutlich, dass keine ITT-Analyse durchgeführt worden war, doch es fehlt die Angabe darüber. Es besteht somit die Möglichkeit, dass die Autoren sich noch nicht mit dem Begriff der ITT auseinandergesetzt hatten. Daher wurde das Kriterium als nicht erfüllt betrachtet.

Wichtig ist auch der Ergebnisbericht der primären und sekundären Zielkriterien und des zu den Ergebnissen passenden Konfidenzintervalls (CONSORT Nummer 17). Dieses sollte nicht für jede Behandlungsgruppe einzeln aufgeführt sein, sondern den Unterschied zwischen den Gruppen deutlich machen, und es kann ergänzt werden durch die Angabe

von p-Werten. Am besten lassen sich Studienergebnisse in einer Tabelle darstellen und durch die Berechnung von NNTs unterstützen (Altman 2001, S.682 [48]). Leider waren in keiner der untersuchten Studien (0%) NNTs und ebenso wenig ARRs, RRRs und Odds Ratios angegeben. P-Werte hingegen fanden sich in allen bis auf drei Veröffentlichungen (92%). Die Ergebnisse der primären Zielkriterien waren in 25 Veröffentlichungen (66%) dargestellt, sieben Studien (18%) machten dazu knappe Angaben oder berichteten diese nur in einem anderen Bericht über die selbe Studie, und in sechs RCTs (16%) fehlten Angaben, ob es sich bei den dargestellten Ergebnissen um den Bericht der primären Zielkriterien handelte (siehe „Primäre Zielkriterien“ in Diagramm 3). Eine Darstellung der Primärresultate einer Studie kann zum Beispiel wie in Noseworthy 1998 [27] (S.1346) erfolgen: „*Effect of sulfasalazine on the primary outcome measure. The primary end point of the study was time to confirmed worsening of the EDSS by 1.0 point. As illustrated in figure 2A, this is a negative trial... . The relative hazard of progression among drug-treated patients was 0.99...*“. Mit diesen Angaben kann der Leser genau erkennen, dass hier über das primäre Zielkriterium berichtet wird. Da die Darstellung der Ergebnisse in dieser Veröffentlichung außerdem ausführlich genug erfolgt, wird das CONSORT-Kriterium als erfüllt bewertet. Eine weitere Möglichkeit, die Primärresultate zu kennzeichnen, ist, wie in Myhr 1999 [25] (S.1050ff.), zunächst im Methodenteil einer Veröffentlichung Primär- und Sekundärkriterien zu definieren und diese dann im Resultateteil abzuarbeiten.

Bei den sekundären Outcome-Kriterien war die Datenlage allgemein etwas ungünstiger. 15 Veröffentlichungen (39%) berichteten darüber, zwölf RCTs (32%) machten zumindest knappe Angaben oder verwiesen auf Schwesterstudien, und in sieben Studien (18%) konnte keine ausreichende Darstellung der sekundären Zielkriterien gefunden werden. Vier Publikationen schließlich (11%) verwendeten laut ihren Angaben bzw. anderweitig aus dem Text zu schließen keine zweitrangigen Outcome-Kriterien (siehe „Sekundäre Zielkriterien“ in Diagramm 3). Auch für Sekundärkriterien dient Noseworthy 1998 [27] (S.1346) als gutes Beispiel. Wegen der Ausführlichkeit der Beschreibungen kann jedoch an dieser Stelle nur der Anfang des Berichts über Sekundärkriterien in dieser Publikation zitiert werden. Für den gesamten Text sei auf die Originalstudie verwiesen. Der Bericht über die Sekundärkriterien wird folgendermaßen eingeleitet: „*Secondary analyses. The protocol specified a number of secondary data analyses as “hypothesis generating“ to identify any beneficial effects. Subgroup analysis: progressive MS. Among progressive patients there was a suggestion of a drug effect as seen in figure 2B...*“ (Noseworthy 1998, S.1346 [27]). Der Leser kann also genau zwischen Primär- und Sekundärkriterien unterscheiden und somit die Wertigkeit der jeweiligen Untersuchungen erkennen.

Als letzter Aspekt der CONSORT-Forderung Nummer 17 sollte ein Konfidenzintervall (CI) genannt sein. Dieses war nur in acht Studien (21%) angegeben, während 30 Veröffentli-

chungen (79%) keine Daten zu einem Konfidenzintervall enthielten (siehe „Konfidenzintervall“ in Diagramm 3). Der Bericht des CI kann zum Beispiel wie in Optic Neuritis Study Group 1997 [29] (S.1407) erfolgen: „... CDMS developed within 5 years in only 25 patients ... and was more likely in women (rate ratio, 2.03; 95%CI, 0.59 to 7.03) and in Caucasians (rate ratio, 2.07; 95%CI, 0.45 to 9.65). However, because of the small number of men and non-Caucasians, the CIs are wide and we cannot be certain of a true association.“ Zusätzlich positiv zu vermerken ist hier, dass gleich im Anschluss an die Nennung der Werte auch die Bedeutung der weiten CIs erklärt wird. Leider wurde jedoch in keiner der in die Bewertung eingeschlossenen Veröffentlichungen zu jedem Outcome-Kriterium das CI angegeben.

Diagramm 10 fasst die Ergebnisse der Untersuchungen zur Angabe von Zielkriterien und des entsprechenden Konfidenzintervalls zusammen und stellt die absoluten Werte für keine, knappe und ausreichende Beschreibung gegenüber.

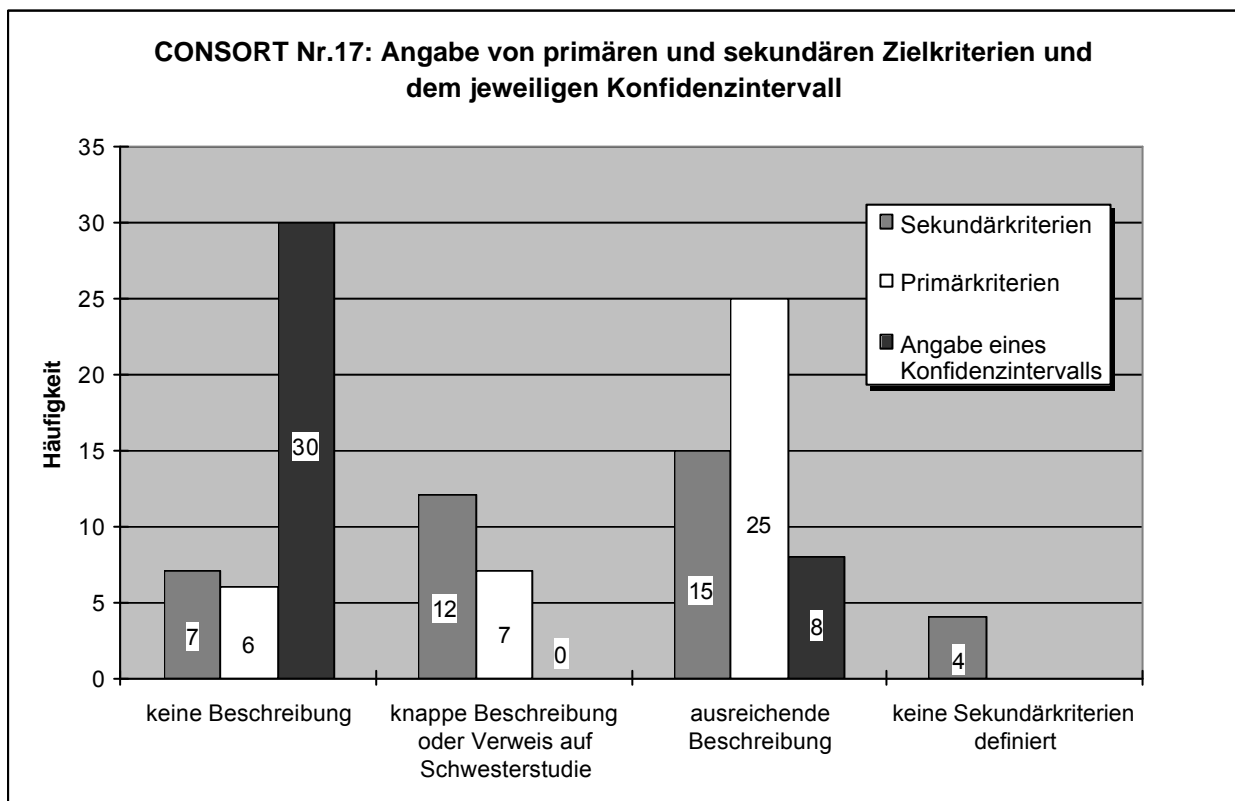


Diagramm 10: Erfüllung von CONSORT-Forderung Nummer 17: Beschreibung von primären und sekundären Zielkriterien inklusive Konfidenzintervall. Die Zahlen in den Balken geben die Anzahl der Studien wieder.

Ein weiterer Aspekt der CONSORT-Kontrollliste ist mit der Nummer 18 die Forderung nach Angabe von weiteren Tests, insbesondere von Subgruppenanalysen, mit einer Erklärung, ob diese Tests schon vor Beginn der Studie geplant waren oder erst im Nachhinein zusätzlich ergänzt wurden. Außerdem sollten alle Angaben dazu komplett aufgelistet werden. Es muss beachtet werden, dass vielfache Analysen die Wahrscheinlichkeit von falsch-positiven Ergebnissen erhöhen (Altman 2001, S.683 [48]). 24 Studien (63%) führ-

ten anscheinend keine zusätzlichen Analysen durch. Von den Studien, die Subgruppenanalysen vollzogen, enthielten fünf (36%; 13% aller Studien) nur knappe Daten oder Verweise auf Nebenstudien, während neun Publikationen (64%; 24% aller Studien) ausreichende Angaben machten (siehe Diagramm 11).

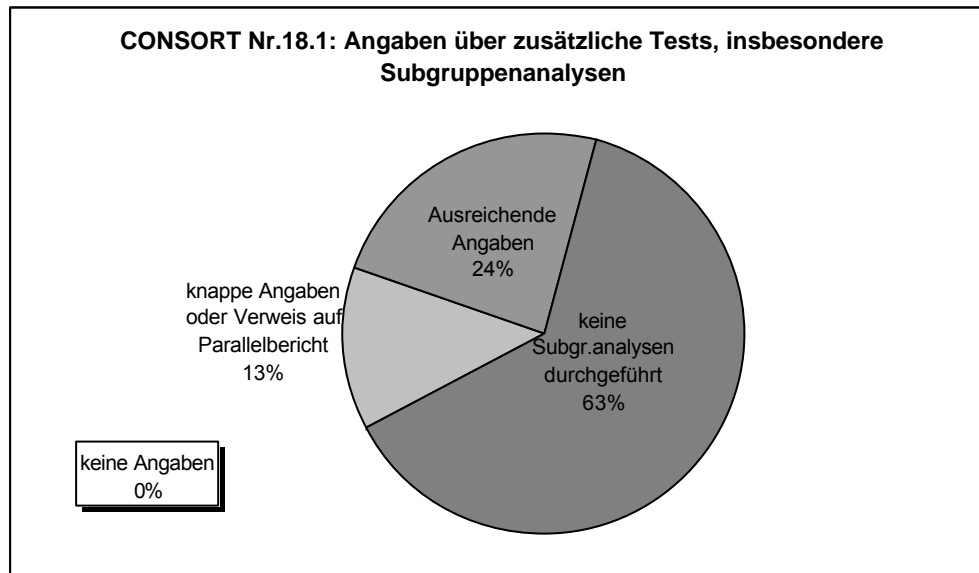


Diagramm 11: Angaben über zusätzliche Tests, insbesondere Subgruppenanalysen.

In nur einer (Li 1999, S.198 [18]) der 14 Veröffentlichungen mit Subgruppenanalysen (7%) war jedoch eindeutig beschrieben, ob diese von Beginn an geplant waren oder erst nachträglich durchgeführt wurden (siehe Diagramm 12).

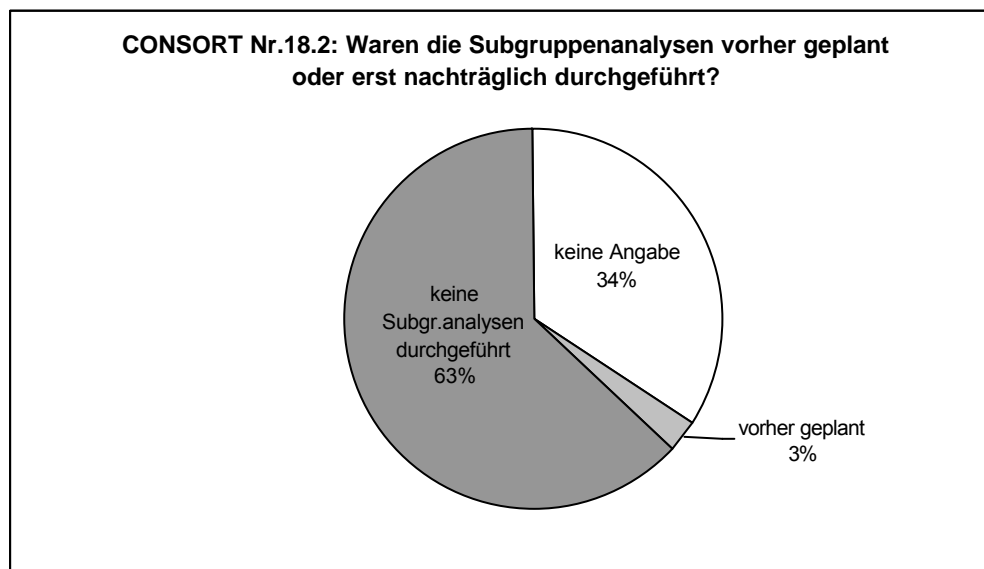


Diagramm 12: Subgruppenanalysen von Beginn an geplant oder nachträglich durchgeführt?

In dieser Studie wird bereits im Methodenteil von zwei Kohorten gesprochen: „COHORT 1. All 560 patients were to undergo five biannual proton density/T2-weighted ... MRI scans COHORT 2. A subgroup of 205 patients from seven centers were to undergo 11 monthly PD/T2 and gadolinium-enhanced/T1-weighted ... MRI scans” (Li 1999, S.198 [18]). Anschließend werden die Resultate im Ergebnisteil eindeutig nach Kohorte 1 und 2

aufgeschlüsselt berichtet (Li 1999, S.200 [18]). Das ist deshalb wichtig, weil der Leser nur so zwischen Haupt- und Nebenuntersuchungen differenzieren kann. Um die CONSORT-Bewertung in diesem Schritt nicht in zu viele Einzelbereiche zu untergliedern, wurde für die vorliegende Arbeit bewusst darauf verzichtet, die Beurteilung nochmals aufzuteilen in Publikationen, in denen Subgruppenanalysen zwar im Methodenteil, aber nicht mehr im Ergebnisteil und umgekehrt berichtet waren. Oftmals musste, wie auch schon an anderer Stelle, die Bewertung teilweise subjektiv erfolgen.

Die letzte Forderung für den Ergebnisteil ist schließlich die Beschreibung aller wichtigen unerwünschten Arzneimittelwirkungen (=UAW), aufgeschlüsselt nach den Behandlungsgruppen und nach deren Häufigkeit. Diese Aufschlüsselung findet oftmals nicht statt, dennoch können in mehr als $\frac{3}{4}$ aller bewerteten Studien die verursachten UAW nachvollzogen werden: 22 Publikationen (58%) gaben die Nebenwirkungen (=NW) direkt an und für weitere sieben Veröffentlichungen (18%) konnten die UAW aus Schwesterstudienberichten ersehen werden. In immerhin neun RCTs (24%) waren keine NW angegeben (siehe „Nebenwirkungen“ in Diagramm 3).

Sehr gut beschrieben sind die Nebenwirkungen in The Once Weekly Interferon for MS Study Group 1999 [42], S.682ff.: “Adverse events associated with IFN therapy are reported in table 5 [siehe Abbildung 16]. Importantly, depression was not increased with therapy, and no injection site necrosis was seen. Thirteen patients experienced 14 serious adverse events: 3 in the placebo group (hernia, vomiting/dehydration, thrombophlebitis), 5 in the low-dose group (cholelithiasis, depression, flu, lymphopenia, trigeminal neuralgia), and 6 in the high-dose group (cholelithiasis [two cases], ovarian mass, pyelonephritis, gastroenteritis, depression with suicide attempt). There were six adverse events (five in the 44µg qw group and one in the 22µg qw group) leading to withdrawal of therapy, including depression in two patients ... and systemic or local side effects of therapy in four patients, all in the 44µg group. ...” Der Leser kann also die wichtigsten Nebenwirkungen direkt aus dem Text den einzelnen Therapiegruppen zuordnen und zusätzlich aus Table 5 (siehe Abbildung 16) eine noch genauere Aufschlüsselung erhalten.

3.1.1.2.4. Ergebnisse der Untersuchung der Diskussion

Für den Diskussionsteil stellt die CONSORT-Gruppe weitere drei Forderungen auf. Nummer 20 der Liste verlangt eine Interpretation mit Berücksichtigung von Studienhypothesen, möglichen Fehlerquellen und Gefahren, die durch eine Vielzahl an Auswertungen entste-

hen können. Dieser Punkt ist nochmals in fünf Unterpunkte gegliedert (Altman 2001, S.683f. [48]):

- 20.1. Kurze Zusammenfassung der Ergebnisse
- 20.2. Mögliche Mechanismen und Erklärungen
- 20.3. Vergleich mit anderen Studien (möglichst in Form eines systematischen Reviews)
- 20.4. Limitationen inkl. Methoden, um diese Limitationen zu minimieren
- 20.5. Zusammenfassung der Auswirkungen

In der Auswertung für Punkt 20 und dessen Unterpunkte fiel zunächst positiv auf, dass mit Ausnahme einer Studie, in der keine Ergebnisdiskussion stattfand und einer Veröffentlichung, in der der Bericht sehr knapp war, alle eingeschlossenen Publikationen eine kurze Zusammenfassung ihrer Ergebnisse (Nr. 20.1) in der Diskussion veröffentlichten. In Teksam 2000 [40] (S.889) findet sich als Zusammenfassung der Resultate lediglich folgende Angabe: "We demonstrated statistically significant decrease in the number of lesions and lesion load with IV IgG, despite the small number of patients." Diese Angaben müssen als eher knapp angesehen werden.

Mögliche Mechanismen und Erklärungen (Nr. 20.2) für die Studienergebnisse wurden in der Hälfte aller Publikationen ausreichend beschrieben, eine knappe Darstellung fand sich in vier RCTs (13%), und fehlende Angaben gab es in 14 Publikationen (37%). Ein gutes Beispiel für eine Beschreibung der möglichen Mechanismen findet sich in Simon 1998 [36] (S.84): „The mechanisms of the therapeutic benefits of β -interferons in MS are incompletely understood, but they undoubtedly involve immunomodulatory actions, possibly by antagonistic interactions with certain proinflammatory molecules, such as IFN- γ . IFN β -1a appears to influence the evolution of MS lesions, including an effect on factors leading to blood-brain barrier disruption, although it is not known if this effect occurs directly or indirectly. It is likely that this mechanism applies to β -Interferons as a class, as a decline in the number of enhancing ... lesions was observed in a majority of patients treated in a pilot study with interferon β -1b ... and with IFN β -1a. ...”.

Ein Vergleich mit anderen Studien zu dem selben Thema (Nr. 20.3) konnte zwar in 26 Veröffentlichungen (69%) ausführlich und in weiteren acht Studien (21%) knapp bestätigt werden, keine führte diesen jedoch in Form eines systematischen Reviews durch. Zwei Publikationen (5%) verglichen ihre Ergebnisse nicht mit anderen Studien, und zwei weitere Studien (5%) waren die ersten klinischen Studien auf diesem Gebiet (The Lenercept Multiple Sclerosis Study Group 1999 [41]; Tan 2000 [39]).

Limitationen inklusive Methoden, um diese Limitationen zu minimieren (Nr. 20.4) wurden in zwölf RCTs (32%) ausreichend beschrieben. 15 Veröffentlichungen (39%) erklärten die Limitationen in relativ knapper Ausführung, oder sie erwähnten nur einen Teil der bei der

Durchführung aufgetretenen Probleme. In elf Studien (29%) fehlten diese Beschreibungen. Eine gute Beschreibung eines Teils der Limitationen findet sich beispielsweise in Millefiorini 1997 [19] (S.158): „... MTX showed no reduction in left ventricular ejection fraction... . These findings have several possible explanations. Firstly, as suggested from previous reports, we studied only young patients who had not been exposed to previous cardiotoxic therapies or mediastinal radiation, and had no pre-existing cardiac pathology, reducing the risk of cardiotoxic effects of MTX Secondly, it should be emphasized that we used low doses of MTX with an intermittent infusion schedule. This pharmacological schedule is different from those used in previous reports where the drug was used to obtain an antineoplastic effect.“ Derartige Angaben sind wichtig, da man so auch über mögliche Ursachen eines Therapieerfolgs- oder misserfolgs aufgeklärt wird, die man sonst nicht erkennen würde. Auch Koudriavtseva 1998 [17] (S.117) beschreibt gut gegliedert die Gründe für eine verminderte Aussagekraft: "There are three reasons why the results of the present trial do not allow us to draw conclusions as to the long term clinical efficacy of mitoxantrone on the course of multiple sclerosis: (1) The relation between short term gadolinium enhancement and long term disability is uncertain; ... (2) The apparent short term clinical benefits we saw were unblinded observations and are therefore not wholly reliable. (3) We cannot exclude the possibility that part of the benefit we saw in the mitoxantrone group came from the addition of methylprednisolone to the treatment regime.“ Eine Zusammenfassung der Auswirkungen der Studienergebnisse auf Klinik und Forschung (Nr. 20.5) war in 27 Studien (71%) zu finden. Neun Publikationen (24%) beschrieben die Auswirkungen nur in einem einzigen Satz oder Nebensatz, und lediglich zwei Veröffentlichungen (5%) machten zu diesem Punkt keine Aussagen. Filippi, J Neurol Sci 2000 [8] (S.44), beschreibt die Auswirkungen zum Beispiel so: "However, our results suggest that the magnitude of such an effect is likely to be clinically unimportant and we conclude that cladribine does not have a major impact on the mechanisms leading to severe tissue destruction in lesions of progressive MS patients.“ Hiermit wird eindeutig die fehlende Möglichkeit eines Einsatzes von Cladribin bei MS und damit die klinische Auswirkung erklärt.

Um eine Gesamtbewertung von Forderung Nummer 20 vornehmen zu können, wurde für die vorliegende Arbeit folgender Schlüssel festgelegt:

„Ausreichende Erfüllung“ wurde zugesprochen, wenn eine Studie

- maximal in einem der fünf Unterpunkte mit „unzureichend“
- oder zweimal mit „zum Teil“

bewertet wurde.

Für „teilweise Erfüllung“ durfte eine Studie

- entweder maximal zweimal „unzureichend“
- oder einmal „unzureichend“ und zweimal „zum Teil“
- oder viermal „zum Teil“

erhalten. Konnte eine Veröffentlichung diese Kriterien nicht erfüllen, wurde sie mit „zu knappe Angaben“ bewertet. Aufgrund dieses Schlüssels wurde Forderung Nummer 20 in 13 Publikationen (34%) ausreichend erfüllt, 20 Veröffentlichungen (53%) enthielten die verlangten Angaben zum Teil, und fünf Studien (13%) konnten dieses Kriterium nicht erfüllen (siehe „Berücksichtigung versch. Kriterien b. Interpretation“ in Diagramm 3).

Als nächstes wird nach einer ausreichenden Beschreibung der Generalisierbarkeit der Studienergebnisse gefragt – Forderung Nummer 21. Wichtig dabei ist, dass die Grundlagen für die Generalisierbarkeit, also die externe Validität, nicht erst in der Diskussion, sondern bereits im Methodenteil einer Studie gelegt werden. Nur so kann die interne Validität bestimmt werden, die Voraussetzung für die externe Validität bzw. Anwendbarkeit einer Studie ist (Altman 2001, S.684 [48]). Wichtigste Angaben im Methodenteil sind dabei die Auswahlkriterien der Patienten mit Studienorganisation- und Ort (CONSORT-Forderung Nummer 3), die Interventionen und deren Applikation (Nr.4), die Definition der Zielkriterien (Nr. 6) und der Zeitraum, in dem die Patienten in die Studie aufgenommen wurden (Nr.14) (Altman 2001, S.685 [48]). Leider konnte eine Beschreibung der Generalisierbarkeit nur in drei Studien (8%) gefunden werden, wobei schon ein Satz wie "We believe our results ... to have applicability to most patients presenting with optic neuritis as a first demyelinating event. ..." (Optic Neuritis Study Group 1997, S.1411 [29]) als positiv gewertet wurde, da sich hier, zusammen mit den oben genannten weiterführenden Angaben, zeigt, dass die Autoren dieser Publikation sich über die Anwendbarkeit Gedanken machten. 21 Publikationen (56%) enthielten zu diesem Punkt nur knappe Angaben, oder es konnten die nötigen Informationen – trotz fehlender expliziter Beschreibung einer Generalisierbarkeit – aus der Diskussion der jeweiligen Studie herausgelesen werden. In fast 2/5 aller eingeschlossenen RCTs (15 Studien, 39%) konnte aus der Diskussion alleine nicht erkannt werden, ob die Ergebnisse generalisierbar sind (siehe „Generalisierbarkeit“ in Diagramm 3).

Der letzte Punkt auf der CONSORT-Kontrollliste ist die Forderung nach einer Interpretation der Ergebnisse im Kontext mit bestehender Evidence. Als ideal wird es angesehen, wenn dies mit Hilfe eines systematischen Reviews oder einer Metaanalyse erfolgt. Falls jedoch noch keine Voruntersuchungen zu dem entsprechenden Thema vorhanden sind, ist dies natürlich nicht möglich. Ein Review sollte unter anderem enthalten, ob die schon

bestehenden Studien ähnlich gemacht wurden oder in ihrer Durchführung unterschiedlich waren. Auch die Ähnlichkeit der Patienten über verschiedene Studien hinweg sollte dort angesprochen werden (Altman 2001, S.685 [48]).

Wie schon bei der Beschreibung der Ergebnisse von Punkt 20.3 der CONSORT-Liste erwähnt, beinhaltet keine der bewerteten Studien ein systematisches Review oder gar eine Metaanalyse. Diese Einschränkung außer Acht gelassen, erfüllten 30 RCTs (79%) das Kriterium ausreichend und erklärten die Ergebnisse im Kontext mit bestehender Forschung. Dies kann auch bedeuten, dass die Autoren eigene oder noch nicht veröffentlichte Forschungsergebnisse mit einfließen lassen. Im Gegensatz dazu wird in CONSORT Nr. 20.3 gefordert, dass lediglich ein Vergleich mit anderen veröffentlichten Studien erfolgen soll. Deshalb stimmen die Zahlen nicht überein mit den Resultaten aus Forderung 20.3. Sieben Publikationen (18%) machten zu bestehender Evidence nur knappe Angaben und lediglich eine (3%; Koch-Henriksen 2000, S.174 [16]) zeigte weder mit schon bestehenden Studien noch mit anderen Forschungsergebnissen einen Zusammenhang (siehe „Interpretation, aktueller Forschungsstand“ in Diagramm 3). Eine Beschreibung im Kontext mit bestehender Forschung könnte zum Beispiel wie in Edan 1997 [4] (S.117) aussehen: „Open pilot studies of mitoxantrone treatment have suggested some therapeutic effect. ... A recent controlled study showed a reduction in relapses, but no significant difference in the EDSS scale or in the development of active MRI lesions. This study was of 13 patients, and the patient group had less active disease ... than in the present study. ...“ Ein Vergleich mit bereits durchgeführten Studien ist deshalb unumgänglich, weil die Aussagekraft einer Studie allein ohne die Resultate früherer RCTs eingeschränkt ist.

3.1.2. Ergebnisse der Bewertung von Therapiestudien zur Multiplen Sklerose nach dem Sackett'schen EBM-Schema

Wie bereits in der Einleitung dieser Arbeit beschrieben war David Sackett der Pionier der EBM. Er entwickelte ein Schema zur Überprüfung der Validität, Wichtigkeit und Anwendbarkeit von klinischen Studien. Da allerdings eine Beurteilung der Validität alleine mithilfe des Sackett'schen EBM-Schemas nur schwer vollkommen zu objektivieren ist und außerdem dieser Bewertungsbogen nicht die unterschiedliche Relevanz einzelner Punkte hervorhebt, wurde, wie in Gliederungspunkt 2.5 dieser Arbeit beschrieben, das sogenannte objektivierte System nach Hassler mit Vergabe von Wertungseinheiten erstellt, das auf dem EBM-Schema nach Sackett beruht. Im folgenden werden die Ergebnisse der Bewertung anhand des neuen Bewertungsschemas besprochen. Eine Verteilung der vergebenen Punkte ist in Diagramm 13 im Anhang und eine Übersicht über alle Bewertungsergebnisse in Tabelle 5 sowie Diagramm 14 dargestellt.

3.1.2.1. Überprüfung der Validität mithilfe des objektiveren Bewertungssystems

Anhand der sechs Schritte des Sackett'schen Bewertungsschemas und des zusätzlich eingefügten Schrittes wurde die Validität der einzelnen Studien beurteilt. Die Auswertung geschah aufgrund der Angaben im Text der eingeschlossen Publikationen und falls nötig von Schwesterstudien bzw. Nebenberichten. Die Resultate der Bewertung sind einschließlich positiver und zu verbessernder Aspekte der untersuchten Studien in den folgenden Abschnitten genauer erläutert.

3.1.2.1.1. Randomisierung

Bei 33 der bewerteten 38 Studien (86%) wurde die Frage, ob eine Randomisation überhaupt durchgeführt wurde, mit drei Punkten bewertet, das heißt die Patienten wurden den Behandlungen mit Sicherheit zufällig zugeordnet. Bei vier Veröffentlichungen (11%) wurde eine Randomisation wahrscheinlich durchgeführt. Diese RCTs erhielten zwei Punkte. Eine Studie (3%) führte die Randomisation fehlerhaft durch, indem die Patienten selbst wählen konnten, ob sie an einer Randomisation teilnahmen oder nicht. Andernfalls wurde ihnen eine Therapie mit bekannter Medikation angeboten (Koch-Henriksen 2000, S.173 [16]). Da aber dennoch randomisiert worden war, ist dieser Studie ein Punkt zuzusprechen (siehe Diagramm 15).

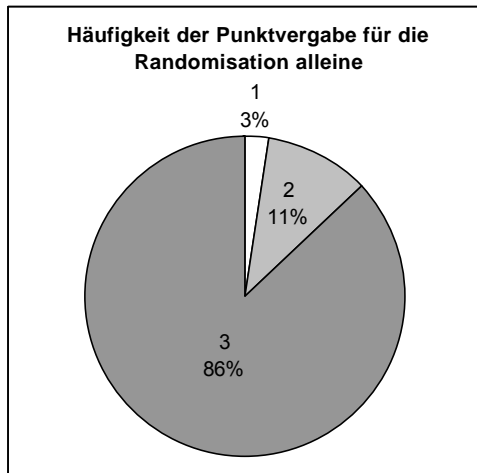


Diagramm 15: Häufigkeit der Punktvergabe für die Randomisation alleine.

Die Ziffern stehen für die vergebenen Punkte und die jeweilige Zahl der eingeschlossenen Studien in Prozent.

Auch der Zeitpunkt der Zufallszuteilung ist wichtig. In 34 Publikationen (90%) wurden die Patienten im richtigen Moment den Behandlungsgruppen zugewiesen, bei zwei Studien (5%) war nicht sicher zu erkennen, wann die Zuteilung im zeitlichen Ablauf der Studie erfolgte, und in zwei Publikationen (5%) wurde der falsche Zeitpunkt für die Randomisierung gewählt. Eine dieser beiden Veröffentlichungen berücksichtigte nach der Zufallszuteilung einen Teil der Patienten nicht, weil diese keine Läsionen im MRT aufwiesen (Koudriavtseva 1998, S.405 [17]) und eine schloss ebenfalls *nach* der Randomisierung nur einen Teil der Patienten in die Studie ein: "MRIs from 342 MS patients with baseline and at least one annual MRI scan were available for analysis and met the standards for inclusion in the study." (Zhao 2000, S.201 [46]) (siehe Diagramm 16).

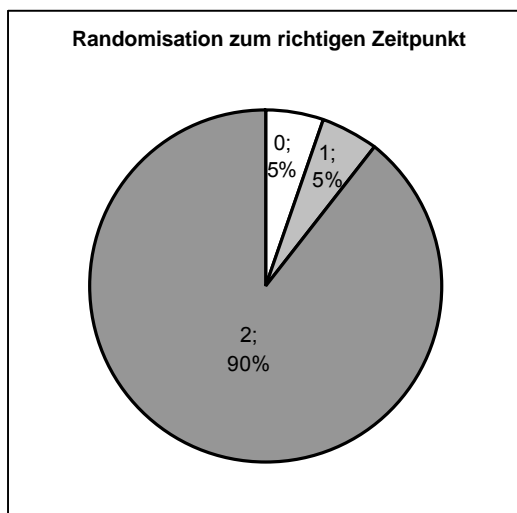


Diagramm 16: Häufigkeit der Punktvergabe für die zum richtigen Zeitpunkt durchgeführte Randomisation.

Die Ziffern stehen für die vergebenen Punkte:
 0 = Randomisation zum falschen Zeitpunkt
 1 = Zeitpunkt der Randomisation nicht eindeutig beschrieben
 2 = Randomisationszeitpunkt korrekt.

Der dritte Punkt in Abschnitt 1 des Bewertungsschemas ist die Geheimhaltung der Zuteilung. Dabei war in lediglich zwölf Publikationen (32%) ein ausreichender Verschluss der Zuteilungsdaten gewährleistet. Eine gute Beschreibung der Geheimhaltung zeigt zum Beispiel Goodkin 2000 [10] (S.1415): "Randomization was performed centrally, with the allocation sequence computer generated and concealed in sequentially numbered, opaque envelopes until the patient was randomly assigned." Auch die Once Weekly Interferon for MS Study Group 1999 [42] bietet ein gutes Beispiel auf Seite 680: "Treatment

assignment was determined by a computer-generated randomization list produced by the Corporate Biometrics Department of Ares-Serono. The randomization code for each patient was delivered to the investigator in sealed envelopes to be opened only in emergency situations requiring knowledge of treatment assignment." Schließlich beschreibt die Lenercept Multiple Sclerosis Study Group nachvollziehbar die Geheimhaltung: "... During the conduct of the study, the randomization list was available only to the Safety Review Board (SRB) members ..." (The Lenercept Multiple Sclerosis Study Group 1999, S.458 [41]). Zu knapp ist eine Beschreibung der Geheimhaltung, wenn nur erwähnt wird, dass die Patienten "centrally randomised" wurden, wie zum Beispiel in Zhao 2000 (S.201 [46]). Dies lässt zwar einen Verschluss der Zuordnungsdaten vermuten, aber nicht nachvollziehen, und war in insgesamt elf Studien (29%) der Fall. Eine fehlende Geheimhaltung oder keine Angaben zu dieser Frage gab es mit 15 Studien in fast zwei Fünftel aller Fälle (39%), zum Beispiel auch in Optic Neuritis Study Group 1997 [29], wobei in den ebenfalls untersuchten Schwesterstudien auch keine genauen Angaben zu finden waren (siehe Diagramm 17). Der Leser wird lediglich informiert, dass „a permuted block design with a separate sequence for each clinical center“ (Beck 1992, S.582 [61]) verwendet worden war, um Patienten „randomly in equal numbers to three treatment groups“ (Beck 1992, S.582 [61]) zu verteilen. Eine Information über eine Geheimhaltung der Randomisationsliste ist hier nicht enthalten.

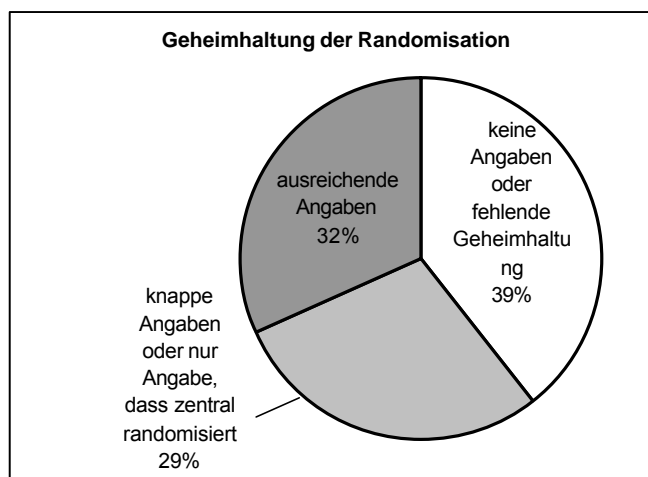


Diagramm 17: Häufigkeit der Punktvergabe für die Geheimhaltung der Randomisation.

Insgesamt erhielten 28 Publikationen (73%) für die Randomisation mit dem objektiven Bewertungsschema sechs oder sieben Punkte. Neun Veröffentlichungen (24%) konnten vier oder fünf Punkte zugesprochen werden, und eine Publikation (3%) erhielt zwei Bewertungspunkte (siehe Diagramm 18). Diese Studie (Koch-Henriksen 2000, S.173 [16]) hätte aufgrund mangelhafter Randomisation auf Validitätsstufe 2 abgewertet werden müssen. Allerdings waren dort die Fehler in den Schritten 3, 4 und 6 des EBM-Schemas größer, sodass eine Abwertung auf Stufe 0 erfolgte.

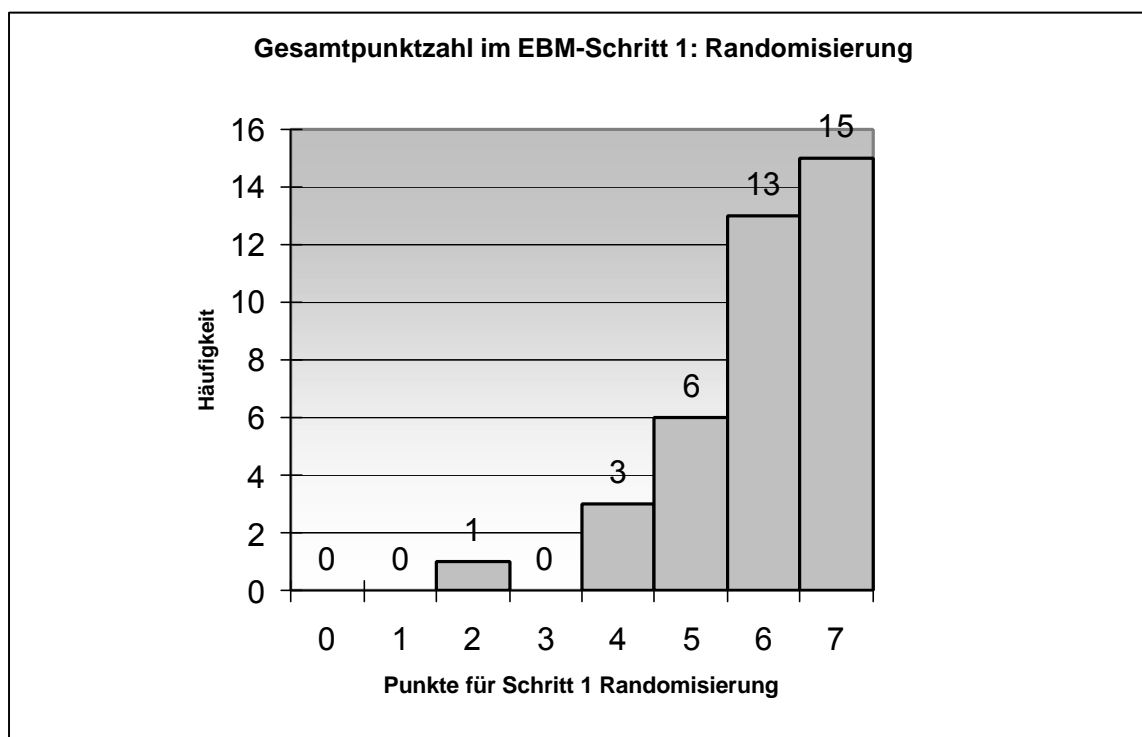


Diagramm 18: Häufigkeit der Gesamtpunktzahl bei der Bewertung der Randomisation. Die Ziffern oberhalb der Balken geben jeweils die Anzahl der Publikationen wieder.

3.1.2.1.2. Anzahl am Ende verbleibender Patienten

Prozentzahl verbliebener Patienten	Häufigkeit (Anzahl Studien)	vergebene Punkte
<70%	7	0
71-75%	1	1
76-80%	1	2
81-85%	6	3
86-90%	4	4
91-95%	7	5
>95%	12	6

Tabelle 10: Verteilung der Anzahl am Ende ausgewerteter Patienten.

Der zweite Schritt in der Auswertung nach Sackett ist die Frage, ob bei der Gesamtbeurteilung einer Studie alle darin aufgenommenen Patienten berücksichtigt werden. Tabelle 10 zeigt die Verteilung der Patientenzahlen in Prozent der am Anfang eingeschlossenen Patienten. Positiv zu vermerken ist dabei, dass genau die Hälfte aller Studien (19 RCTs) mehr als 90% der anfangs aufgenommenen Patienten am Ende aus-

wertet und damit fünf oder sechs Bewertungspunkte erhält. Ein Beispiel für ein gutes Bemühen der Autoren, möglichst viele der anfangs eingeschlossenen Patienten in die Endauswertung einzubeziehen, zeigt Miller 1999 [20], wo Daten bis zum letzten Abschluss der Studie evaluiert wurden, „including all data collected after the time of cutoff for the interim analysis, which lead to early termination of the study“ (Miller 1999, S.852 [20]).

Ein gutes Viertel der Studien (27%) zog zwischen 81 und 90% der Studienteilnehmer für die Beurteilung der Zielkriterien heran (3 bis 4 Punkte), und zwei Publikationen liegen zwischen 71 und 80% (1 bis 2 Punkte). Sieben Veröffentlichungen (18%) werteten weniger

als 70% aller anfangs eingeschlossenen Patienten aus und erhielten somit 0 Punkte, da diese Studien dann nicht mehr als valide gelten können (siehe Diagramm 19 und Tabelle 5).

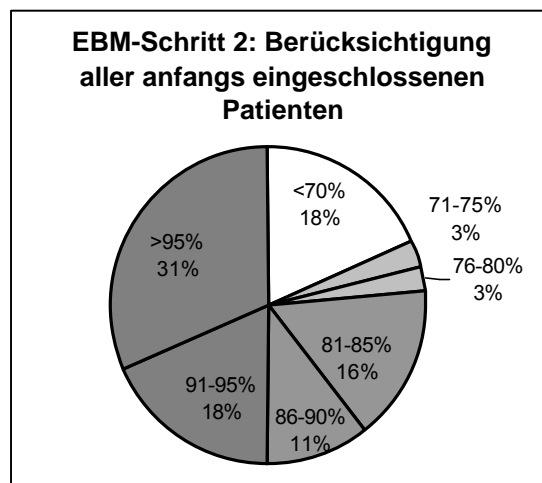


Diagramm 19: EBM-Schritt 2: Berücksichtigung aller anfangs eingeschlossenen Patienten.

Diese sieben Publikationen mussten infolgedessen auf Validitätsstufe 0 abgewertet werden. Zwei Studien (Simon 1998 [36] und Sørensen 2000 [38]) hätten eine Abwertung auf Validitätsstufe 2 erhalten. Eine war jedoch aufgrund ihrer Gesamtpunktzahl schon in dieser Stufe eingeordnet (Simon 1998 [36]) und eine konnte mit Stufe 1 sogar nur eine niedrigere Wertung (Sørensen 2000 [38]) erhalten. Ebenfalls ungünstig ist, wenn innerhalb einer Publikation oder auch in zwei verschiedenen Veröffentlichungen zu einer Studie unterschiedliche Angaben die Patientenzahl betreffend gemacht werden. So fällt zum Beispiel in der Studie der Optic Neuritis Study Group auf, dass in der bewerteten Publikation (Optic Neuritis Study Group 1997, S. 1405f. [29]) von 388 Patienten berichtet wird, während in Beck 1993 389 Studienteilnehmer angegeben werden (Beck 1993, S.1765 [62]). Ähnlich ist die Problematik in Koudriavtseva 1998 [17], S.405. Dort werden auf einer Seite zwei unterschiedliche Zahlen an ausgeschlossenen Patienten genannt. Unter „Baseline and follow-up ... characteristics“ findet sich „Eighteen patients ... participating ... were not considered in analysis because of low Gd-enhanced activity“ (Koudriavtseva 1998, S.405 [17]), während unter „Discussion“ auffällt, dass 19 Patienten ausgeschlossen waren: „... 19 patients of the original cohort were ‘less informative’ for evaluation of individual response as they have only few enhancing lesions ...“ (Koudriavtseva 1998, S.405 [17]). Selbst wenn, wie in diesem Fall, an anderer Stelle von einem Patienten berichtet wird, der aus der Therapie ausschied, wodurch sich unter Umständen die Diskrepanz in der Anzahl erklären könnte, bleibt die Gesamtkonstellation dennoch unklar. Dies zeigt wiederum, dass mangelnde Transparenz die objektive Beurteilung einer Studie ungleich erschwert. Ein weiteres Exempel für unterschiedliche Angaben bietet schließlich Myhr 1999 [25]. Dort sind in Tabelle 2 (S.1051) für beide Behandlungsgruppen Patientenzahlen aufgeführt, die sich von den im Text genannten um eins unterscheiden. Zusätzlich entspricht in der IFN-

Hochdosis-Gruppe die in Tabelle 2 genannte Prozentzahl an Studienteilnehmern, die keine neuen Läsionen aufwiesen (65,5%), einer Gesamtzahl von 29 Patienten im Vergleich zu im Text genannten 27 Teilnehmern (S.1051f.) und zu in der Tabelle 2 aufgeführten 28 Patienten. Derartige Diskrepanzen machen es für den Leser unmöglich, die wahren Ergebnisse einer Studie nachzuvollziehen.

3.1.2.1.3. Intention-to-treat-Analyse

Schwierig ist oftmals die Frage nach der Intention-to-treat-Analyse (=ITT) zu beantworten. Denn wenn eine ITT im Text einer Veröffentlichung genannt wird, bedeutet das noch nicht, dass sie auch wirklich durchgeführt wurde. Deswegen werden hier die Angabe und Durchführung getrennt dargestellt. In 22 Publikationen (58%) konnten die Worte „intention-to-treat“ gefunden werden, 16 Studien (42%) enthielten sie nicht (siehe Diagramm 20).

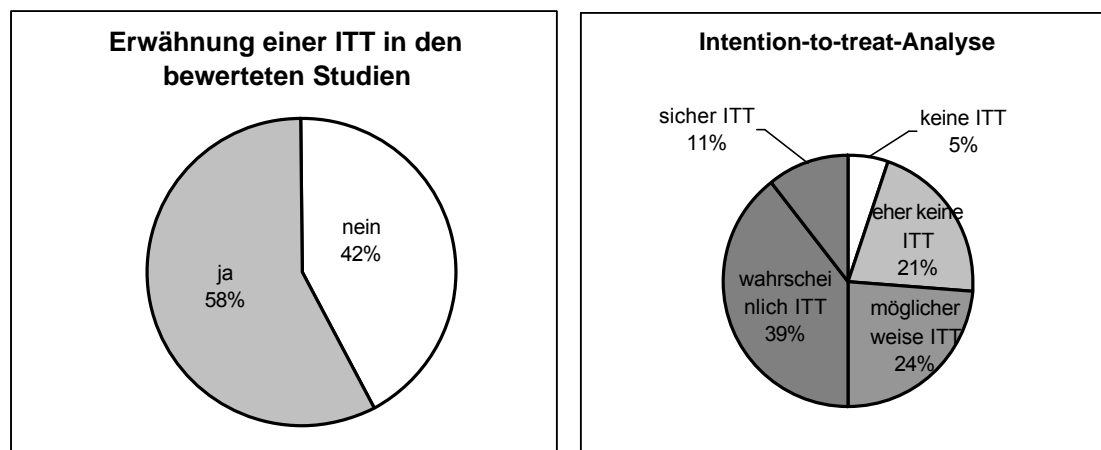


Diagramme 20 und 21: EBM-Schritt 3: Häufigkeit der Auswertung nach ITT.

Die Durchführung betreffend war die Auswertung nur in vier RCTs (11%) mit Sicherheit nach intention-to-treat erfolgt, wofür im EBM-Schema vier Punkte vergeben wurden. Allerdings hatte in 15 Studien (39%), wenn auch nicht ganz eindeutig, eine ITT wahrscheinlich stattgefunden, was mit drei Punkten bewertet wurde. Die Zahlen für „sicher oder wahrscheinlich ITT“ stimmen nicht überein mit denen für die Nennung einer ITT im Text, da in vier Studien trotz Erwähnung eher keine ITT durchgeführt worden war und im Gegenzug nur in einer Veröffentlichung (Patti 1999 [31]) trotz fehlender Erwähnung eine ITT eher trotzdem stattgefunden hatte. Ein gutes Beispiel für eine Auswertung nach ITT zeigt Miller 1999 [20]. Dort wurde in einem teilnehmenden Studienzentrum versehentlich eine von der angegebenen Konzentration abweichende Dosis Kontrastmittel gegeben. „The analyses were therefore performed with and without this site included.“ (Miller 1999, S.851 [20]) Das bedeutet, dass sowohl eine ITT als auch eine Per-protocol-Analyse stattgefunden

den hatte, da auch eine unterschiedliche KM-Gabe unter das Stichwort „Behandlung eines Patienten“ gezählt werden kann.

Neun Publikationen (24%) führten möglicherweise eine ITT durch, wofür zwei Punkte vergeben wurden. Eher keine ITT konnte in acht RCTs (21%) gefunden werden und sicher keine ITT in zwei Studien (5%) (siehe Diagramm 21). Eindeutig ist die Beurteilung dann, wenn in einer Publikation beschrieben steht, dass die Auswertung „not as an `intention to treat study“ (Tan 2000, S.100 [39]) oder dass sie nur „per protocol“ durchgeführt wurde. Aber auch aus Angaben wie zum Beispiel „Both patients started a new treatment course and were excluded from the analysis“ (Oliveri 1998, S.1835 [28]) oder „failure to take scheduled doses of trial medication“ (Myhr 1999, S.1051 [25]) als Ausschlusskriterium lässt sich schließen, dass eine ITT in diesem Fall nicht stattfand.

14 Studien erfüllten die Kriterien, um aufgrund mangelnder oder fehlender ITT abgewertet zu werden. Allerdings erhielt nur eine tatsächlich eine Abwertung mangels ITT, und zwei RCTs erhielten schon aufgrund ihrer Punktzahl die entsprechende Validitätsstufe 2. Die Validität von sieben Publikationen war aufgrund der Gesamtpunktzahl sogar schlechter als die Stufe, die durch eine Abwertung erreicht worden wäre, und in vier Veröffentlichungen waren andere Kriterien vorhanden, die zu einer Abwertung auf Stufe 0 statt Stufe 2 führten (siehe im Anhang in Tabelle 5).

3.1.2.1.4. Verblindung

Wichtig bei der Verblindung ist, zwischen Patienten, Diagnostikern und Therapeuten zu unterscheiden. Dabei fällt auf, dass Therapeuten signifikant seltener als Patienten ($p < 0,01$) oder Diagnostiker ($p < 0,01$) verblindet sind (siehe Diagramm 25): In nur acht Studien (21%) konnte eine volle Verblindung festgestellt und damit zwei Punkte in der Bewertung vergeben werden. 16 Studien (42%) machten Angaben, die auf eine teilweise Verblindung der Therapeuten schließen lassen (1 Punkt) und in 14 RCTs (37%) waren die behandelnden Ärzte nicht verblindet, wofür 0 Punkte zugeteilt wurden (siehe Diagramm 23). Patienten und Diagnostiker hielten sich betreffs der Verblindung über die zugewiesene Therapie etwa die Waage: Diagnostiker waren in 65% (25 Studien) voll verblindet, Patienten in 63% (24 Studien). Dafür gab es je zwei Bewertungspunkte. Eine teilweise Verblindung fand bei diagnostisch tätigen Ärzten in zwölf RCTs (32%) statt und bei Patienten in neun Studien (24%) (je 1 Punkt). Eine fehlende Verblindung gab es bei Diagnostikern in 1 Publikation (3%) und in der Gruppe der Patienten in fünf Veröffentlichungen (13%), wofür kein Punkt eingetragen wurde (siehe Diagramme 22 und 24).

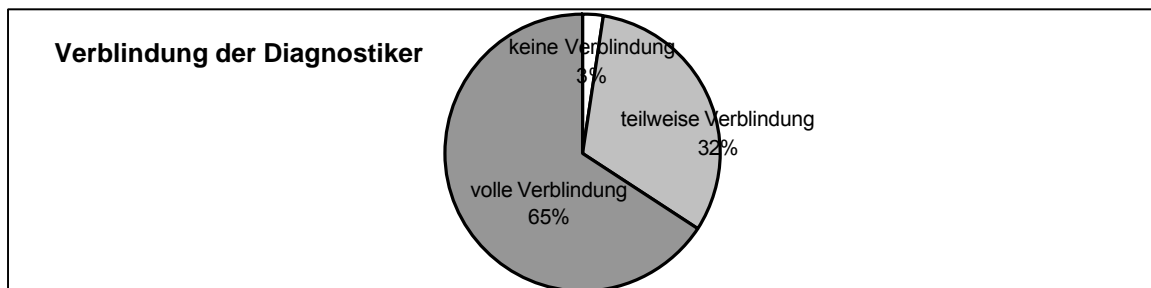
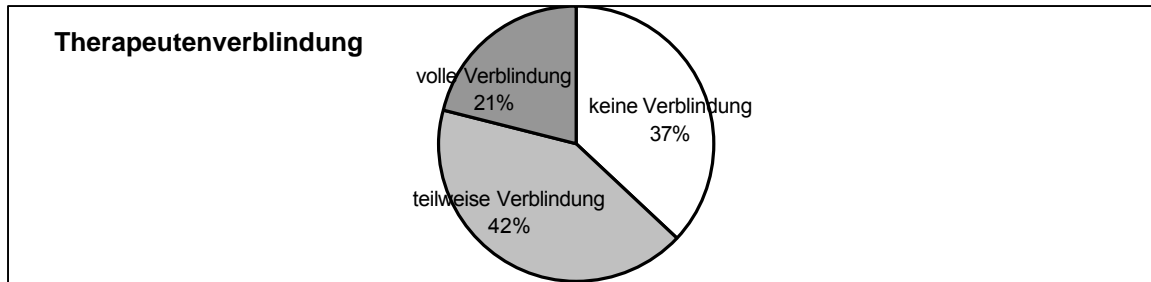
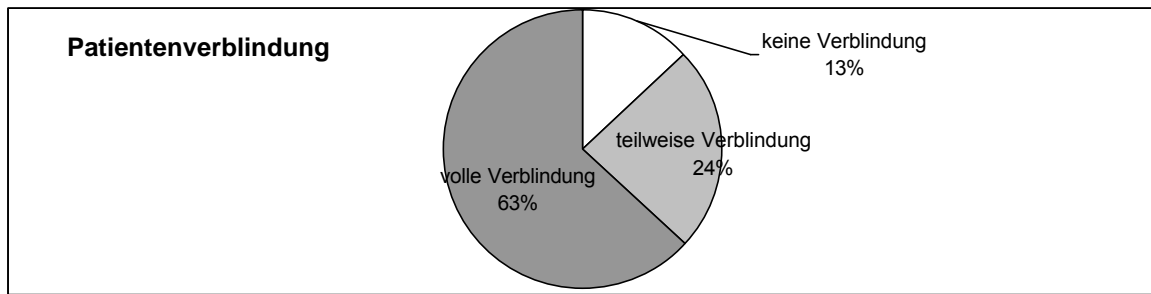


Diagramme 22, 23 und 24: EBM-Schritt 4: Verblindung von Patienten, Therapeuten und Diagnostikern.

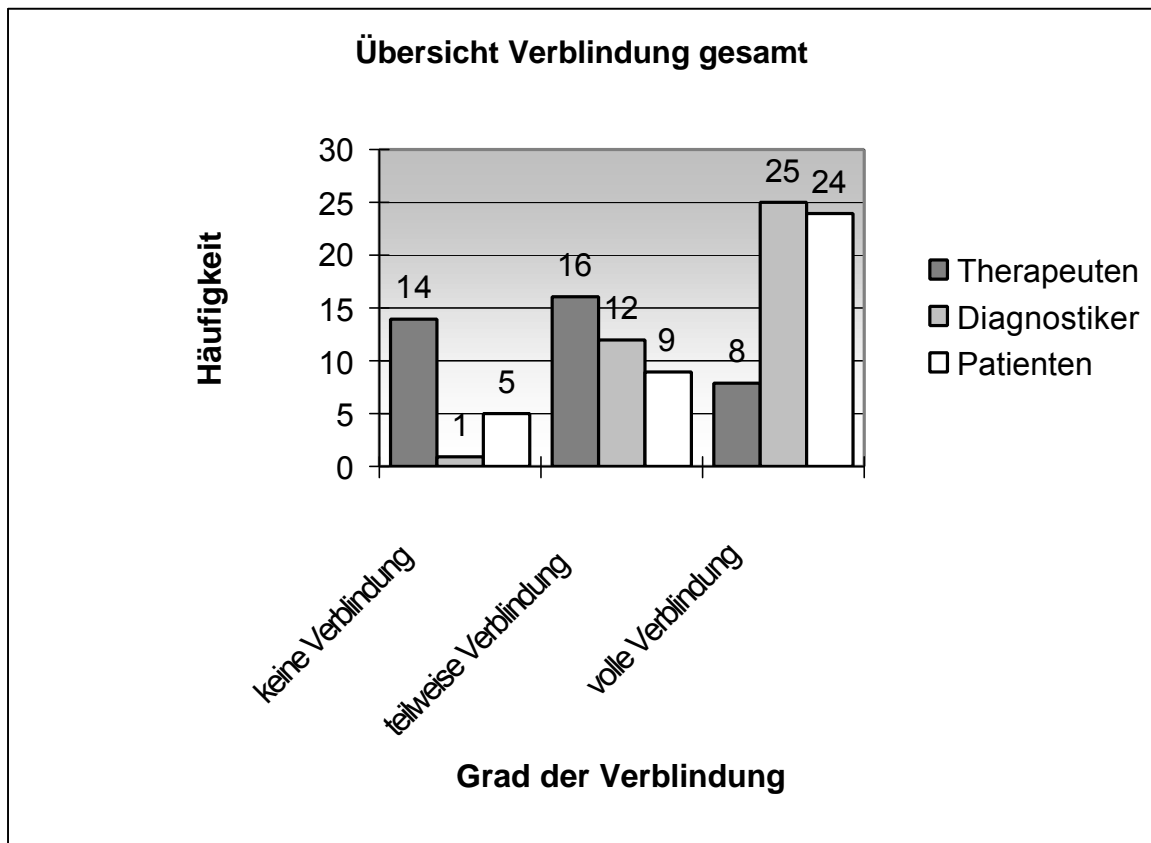


Diagramm 25: EBM-Schritt 4: Vergleich der drei Bereiche der Verblindung.

Gute Angaben für eine gesicherte Verblindung bietet zum Beispiel Oliveri 1998 [28] (S.1274). Diese sind in Gliederungspunkt 3.1.1.2.2. auf Seite 61 der vorliegenden Arbeit zitiert. Sehr transparent und damit bewertungsfreundlich ist es, wenn in einer Publikation erwähnt wird, dass die Verblindung von Ärzten und Patienten untersucht wurde und die Ergebnisse dieser Evaluation in einer Tabelle aufgelistet werden. Auch in van Oosten 1997 [44], S.355 findet sich eine derartige Tabelle. Leider zeigt sich dort jedoch eine fehlende Verblindung der an der Studie teilnehmenden Ärzte. Der Bewerter steht in diesem Fall vor dem Problem, dass eine fehlende Verblindung ansich negativ ist, der Publikation die hohe Transparenz jedoch zugute gehalten werden müsste. In einer EBM-Bewertung gibt es aber keine andere Möglichkeit als die Frage der Verblindung der Diagnostiker in diesem Fall mit „nein“ zu beantworten. Da aber Veröffentlichungen, die viele derartige Angaben machen, auch wenn sie für eine Bewertung negative Auswirkungen haben könnten, nicht schlechter gestellt sein dürfen als Publikationen, die den Leser im Unklaren lassen, müssen alle nicht genannten Fakten als „nicht durchgeführt“ oder „nicht vorhanden“ gewertet werden. Ein weiteres Exempel für eine weniger gute Verblindung zeigt Beck 1993 [62], ein Schwesterbericht von Optic Neuritis Study Group 1997 [29], wo beschrieben wird, dass eine der Behandlungsgruppen nicht verblindet war. Dieser Publikation ist jedoch wiederum zugute zu halten, dass die fehlende Verblindung, im Gegensatz zu Optic Neuritis Study Group 1997 [29], wenigstens erwähnt wird. Ein Zitat der Beschreibung findet sich in Abschnitt 3.1.1.2.2. auf Seite 61 dieser Arbeit.

Neun Publikationen erfüllten die Kriterien für eine Abwertung, davon drei Veröffentlichungen für Validitätsstufe 0 („nicht valide“) und sechs Publikationen für Stufe 2 („eingeschränkt valide“). Von den sechs Stufe-2-Studien waren jedoch vier im Endergebnis allein aufgrund ihrer Punktzahl niedriger eingestuft als die Abwertungsstufe 2 und eine wurde aufgrund mangelnden Patienteneinschlusses sowie fehlenden Vergleichs der Ähnlichkeiten der Vergleichsgruppen als „nicht valide“ eingeordnet. Von den übrigen wegen nicht vorhandener Verblindung oder fehlender Angaben auf Stufe 0 abgewerteten Studien waren zwei allein aufgrund ihrer erreichten Punktzahl nicht valide (Koch-Henriksen 2000 [16]; Koudriavtseva 1998 [17]) und lediglich eine wurde wirklich mangels Verblindung abgewertet (Edan 1997 [4]) (siehe Tabelle 5).

3.1.2.1.5. Gleichbehandlung

Als nächstes Kriterium wurde im Schritt fünf des EBM-Schemas eine Gleichbehandlung von Therapie- und Placebo-Gruppen untersucht. Keine der eingeschlossenen Studien machte dazu eine explizite Aussage, und in den meisten RCTs war die zusätzliche Gabe

von Steroiden bei Exazerbationen erlaubt. Da aber eine der Behandlungsgruppen unter häufigeren Exazerbationen hätte leiden können und Steroide unter Umständen Auswirkungen auf die Verhinderung zukünftiger Rückfälle positiv wirken könnten, ist hier eine Gleichbehandlung nur noch teilweise gegeben. Man kann aber davon ausgehen, dass Studienpatienten außer der experimentellen Intervention mit einer gewissen Wahrscheinlichkeit gleich behandelt werden, wenn die Zuteilung randomisiert erfolgte, eine Verblindung vorlag und die Umgebungsbedingungen für die Vergleichsgruppen identisch sind. Deshalb wurden alle Publikationen, in denen eine eindeutige Aussage fehlte, aber auch nicht das Gegenteil nachgewiesen werden konnte, mit 3 von 5 Punkten für „Gleichbehandlung eher gegeben“ bewertet. Dies war in 36 Publikationen (95%) der Fall. Zwei Studien (5%) behandelten die Vergleichsgruppen eher nicht bzw. zu einem Teil nicht gleich (siehe Diagramm 26).

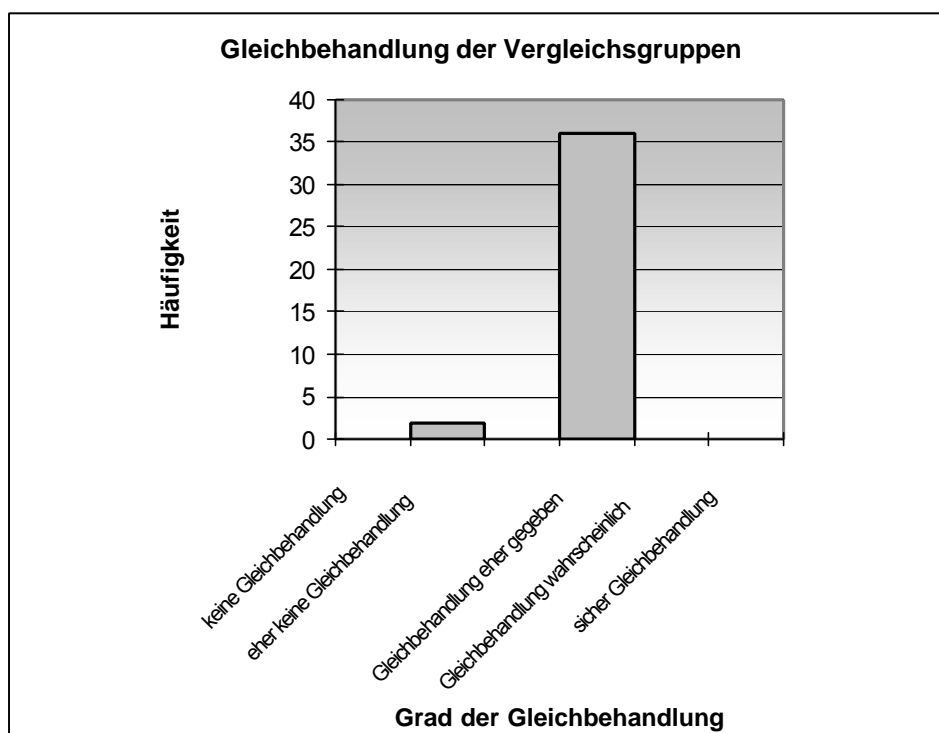


Diagramm 26: EBM-Schritt 5: Häufigkeit der Gleichbehandlung in den Therapiegruppen eingeschlossener Studien.

In der einen Studie gab es zwar für die Gruppe, die orale Kortikosteroide erhielt, eine Orales-Placebo-Gruppe, die jedoch auch als Vergleich zu einer intravenösen Therapie hergenommen wurde (Optic Neuritis Study Group 1997, S.1405 [29]). Somit waren Placebo-Gruppe und I.v.-Therapie-Gruppe nicht gleich behandelt, da die Applikationsform unterschiedlich war. In der anderen Studie war die Häufigkeit der Applikation in den Vergleichsgruppen unterschiedlich: IFN β -1b wurde jeden zweiten Tag verabreicht und IFN β -1a nur einmal pro Woche, jeweils in einer vorher untersuchten Dosierung (Koch-Henriksen 2000, S.173 [16]). Damit lässt sich nicht nachweisen, ob ein möglicher Therapieerfolg auf die Art des Interferons oder auf die Anzahl der Applikationen im Sinne

pieerfolg auf die Art des Interferons oder auf die Anzahl der Applikationen im Sinne eines Placebo-Begleiteffekts zurückzuführen ist. Eine Gleichbehandlung ist dann nur noch sehr eingeschränkt vorhanden. Beide hier zitierte Studien erhielten in der Bewertung nur einen Punkt für „Gleichbehandlung eher nicht gegeben“ und hätten deshalb auf Validitätsstufe 2 abgewertet werden müssen, die sie jedoch aufgrund ihrer Punktzahl unabhängig von einer möglichen Abwertung schon unterschritten hatten.

3.1.2.1.6. Ähnlichkeit der Behandlungsgruppen

Der letzte Schritt der herkömmlichen EBM-Bewertung nach Sackett ist die Untersuchung, ob die Patienten in den Vergleichsgruppen ähnlich waren. Am besten dazu geeignet ist eine Tabelle, normalerweise „Table 1“. Deshalb wurde für die vorliegende Arbeit zunächst überprüft, ob die eingeschlossenen Publikationen oder deren Schwesterstudien Informationen über die Patienten aufgeschlüsselt nach Behandlungsgruppen in Form einer „Table 1“ enthielten. In 26 Veröffentlichungen (68%) war eine derartige Tabelle zu finden, während 12 Publikationen keine Angaben zu den eingeschlossenen Patienten mithilfe einer „Table 1“ machten (siehe Diagramm 27).



Diagramm 27: EBM-Schritt 6/1: Vorhandensein einer Tabelle mit den relevanten Patientendaten.

Die Wichtigkeit einer Patientendatentabelle wird deutlich anhand der gefundenen Korrelationen zwischen vorhandener Auflistung und guter Übereinstimmung der Patientengruppen: 20 Publikationen (71%) von denjenigen, in denen eine Table 1 enthalten war, zeigten weitgehende bis sehr gute Übereinstimmung der Patientengruppen (3 bis 5 Bewertungspunkte), während nur acht Veröffentlichungen (29%) mit vorhandener Patiententabelle eine nur teilweise bis fehlende Ähnlichkeit aufwiesen. Äußerst beeindruckend war die negative Korrelation: Alle zehn Studien (100%), die keine derartige Auflistung enthielten,

konnten keine Übereinstimmung der Patientenmerkmale zeigen, sei es wegen tatsächlich fehlender Ähnlichkeit oder, wie in den meisten Fällen, mangels Angaben.

Bei den Ähnlichkeitsuntersuchungen wurden in insgesamt zehn RCTs (26%) die Vergleichsgruppen für sehr ähnlich oder ähnlich mit Ausnahmen befunden, wofür 5 bzw. 4 Bewertungspunkte vergeben wurden. Weitgehend ähnlich oder zum Teil ähnlich waren

Bewertungseinheiten	Häufigkeit
0	5
1	5
2	8
3	10
4	7
5	3

Tabelle 11: Häufigkeit der vergebenen Punkte in EBM-Schritt 6 (siehe auch Diagramm 28).

18 Studien (48%), das heißt fast die Hälfte aller eingeschlossenen Veröffentlichungen (3 bzw. 2 Punkte). Eine Ähnlichkeit war eher nicht oder überhaupt nicht gegeben in 10 Studien (26%). Das bedeutet, dass hier die Beschreibung fehlte und diese Bewertung aufgrund mangelnder Angaben zustande kam oder dass eine Übereinstimmung tatsächlich nicht gegeben war. Die entsprechenden Veröffentlichungen erhielten einen oder

null Bewertungseinheiten (siehe Diagramm 28). Zu den Bewertungsergebnissen im Schritt 6 siehe auch Tabelle 11.

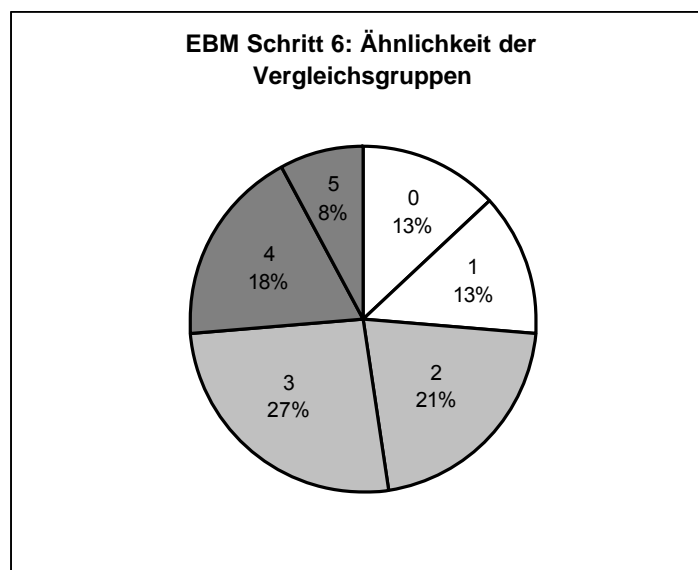


Diagramm 28: EBM-Schritt 6/2: Ähnlichkeit der Vergleichsgruppen. Die Ziffern stehen für die vergebenen Punkte, von 0 = keine Ähnlichkeit der Gruppen bis 5 = Vergleichsgruppen sehr ähnlich; die Prozentzahlen stehen für die Häufigkeit innerhalb der untersuchten Studien (siehe auch Tabelle 10).

Als Beispiel ist die Studie von Teksam und Kollegen 2000 [40] zu nennen, in der keine Table 1 vorhanden ist und auch sonst bis auf Alter und Geschlecht der Studienteilnehmer in den Vergleichsgruppen keinerlei verwertbare Angaben zu den Patienten vorliegen (Teksam 2000, S.887 [40]). Auch Tan 2000 [39] macht keine nützlichen Angaben zu den Patienten. Hier wird nicht einmal Alter und Geschlecht nach den verschiedenen Behandlungsgruppen aufgeteilt, und eine Beurteilung der Gruppenähnlichkeit ist nicht im Entferntesten möglich (Tan 2000, S.99ff. [39]). Negativ ist auch, wenn die Gruppengröße stark differiert, was in Koudriavtseva 1998 1998 der Fall ist. Dort enthält die LD-IFN-Behandlungsgruppe 45% mehr Patienten als die HD-IFN-Gruppe (Verhältnis 29:20), ganz unabhängig davon, dass für diese Veröffentlichung die Randomisation zu einem falschen Zeitpunkt stattfand, indem von den randomisierten Patienten ein Teil nachträglich aus der

Bewertung ausgeschlossen wurde (Koudriavtseva 1998, S.405 [17] und Pozzilli 1996, S.253 [127]).

Im Gegensatz dazu ist ein besonders gutes Beispiel für Gruppenähnlichkeit zum Einen die Studie von Jacobs und Kollegen aus dem New England Journal of Medicine 2000 [13], wo in Table 1 die Patienten sogar zusätzlich nach Rasse, Volumen der Läsionen oder der MS-Familienanamnese getrennt aufgelistet sind und sich zeigt, dass die Behandlungsgruppen in der Tat weitgehende Ähnlichkeit aufwiesen (Jacobs 2000, S.901 [13]). Neben vielen weiteren möglichen positiven Beispielen (sh. Tabelle 5) sei auf die Publikation der Optic Neuritis Study Group 1997 [29] hingewiesen. Die Ähnlichkeit der Vergleichsgruppen konnte dort sehr gut aus Table 1 in Beck 1993 [62], S.1765, ermittelt werden.

Wegen nicht vorhandener oder zu geringer Übereinstimmung der Behandlungsgruppen erfüllten neun Studien die Kriterien für eine Abwertung, von denen allerdings keine durch die Herabstufung selbst in eine niedrigere Validitätsstufe eingeordnet wurde: Fünf Studien (56%) waren allein aufgrund ihrer Punktzahl in die Validitätsstufe 0 eingeordnet worden, und je zwei Studien (22%) waren durch andere Kriterien schlechter gestellt bzw. aufgrund ihrer Punktzahl niedriger eingestuft als in der Stufe 2, in die sie aufgrund der Abwertung eingeordnet gewesen wären.

3.1.2.1.7. Sonstige die Validität beeinflussende Faktoren

Ein von Porzsolt und Kollegen erst nachträglich eingefügter, aber ebenfalls wichtiger Schritt in der EBM-Beurteilung von Studien ist die Frage, ob es darin weitere Faktoren gibt, die die Validität beeinflussen könnten (Porzsolt 2001, S.26 [125]). Die Bewertung dieses Schrittes ist besonders subjektiv, da die Gründe für eine potentielle Abwertung allein aufgrund der Menge an möglichen Fehlern oder an Aspekten mit mangelnder Berichterstattung, aber auch wegen unterschiedlicher experimenteller Therapien, besonders vielfältig sind. 32 Studien (84%) enthielten größere oder kleinere Mängel, die zu einer Abwertung hätten führen können. Da aber alle der betroffenen RCTs schon vor Schritt **C** als nur eingeschränkt oder nicht valide bewertet worden waren, kam es in keiner der untersuchten Publikationen zu einer weiteren Herabstufung der Validität.

Die in diesem Gliederungspunkt im Folgenden angeführten Kritikpunkte sind nur eine Auswahl an Beispielen, die die Validität einer Therapiestudie zusätzlich einschränken könnten. Um den Rahmen dieser Arbeit nicht zu sprengen, wurden die wichtigsten und besten Beispiele ausgesucht und auf eine vollständige Darstellung aller gefundenen Schwierigkeiten verzichtet.

Einer der Fehler, die zu einer Abwertung hätten führen können, waren Schlussfolgerungen, die nicht belegt wurden („data not shown“) und damit für den kritischen Leser nicht nachvollziehbar waren, wie zum Beispiel in Coles 1999 [2] (S.298), Wolinsky 2000 [45] (S.1735), Goodkin 2000 [10] (S.1417) oder Rice 2000 [32] (S.1150).

Eine weitere Bias-Möglichkeit ist ein sogenannter „Conflict of Interest“. Das bedeutet, dass ein Arzt, der an einer Studie mitarbeitet, möglicherweise von der Pharmakologiefirma, deren Produkt untersucht werden soll, Zahlungen erhält oder dass eine Studie generell von einer Arzneimittelfirma finanziert wird. Dies könnte zu verfälschten Ergebnissen führen, da sich die durchführenden Ärzte und andere Verantwortliche dieser Firma verpflichtet fühlen könnten und sich damit in einem Interessenkonflikt befänden. Einerseits möchten zuständige Ärzte in der Regel eine weitgehend objektive Beurteilung eines Medikaments erreichen, andererseits soll dem untersuchten Medikament der Firma zuliebe eine möglichst gute Wirkung nachgewiesen werden. Ein derartiger Fehler wäre denkbar in Jacobs, N Engl J Med 2000 [13] (S.903), oder in Jacobs, J Neuroimmunol 2000 [12] (S.172). Es muss jedoch an dieser Stelle betont werden, dass eine Studie, nur weil sie von der Industrie unterstützt wird, keinesfalls von vornherein als nicht valide betrachtet werden muss. Nur sollten derartige Studien mit besonderer Sorgfalt auf ihre Validität untersucht werden. Ein gutes Beispiel für eine RCT, in der einem Medikament trotz Unterstützung durch eine Pharmakologiefirma keine positive Wirkung nachgewiesen werden kann, zeigt Filippi 2000 [8] (S.44). Der untersuchte Stoff Cladribin zeigt darin keine deutliche Beeinflussung der MS.

Ein weiteres die Validität beeinflussendes Problem ist wie in Sørensen 2000 [38] der gemeinsame Bericht über zwei Studien in einer Publikation wobei das Vorgehen bei der Behandlung der Patienten in beiden Studien unterschiedlich war und die Ausführlichkeit der Angaben über Patienten, Therapie- und Diagnostikmethoden stark differiert, denn der dänische Teil der Publikation ist wesentlich ausführlicher dargestellt (Sørensen 2000 [38]). Das führt dazu, dass der Leser die Validität beider Einzelstudien nicht genau beurteilen kann und somit auch die gesamte Validität der Publikation nur schwierig zu ermitteln ist.

Ungünstig für die Validität einer Studie ist auch, wenn der Leser mit falschen oder an verschiedenen Stellen einer Publikation nicht übereinstimmenden Zahlen konfrontiert wird. So sind zum Beispiel für das selbe Patientenkollektiv in Myhr 1999 [25] (S.1051), Tabelle 2, unterschiedliche Zahlen genannt, und die Prozentwerte passen zum Teil nicht zu den genannten Patientenzahlen. Für die 4.5MIU IFN-a2a-Gruppe ist dabei im Text von 28 Patienten die Rede, wozu auch der Prozentwert von 53,6% für 15 Patienten passen würde. In Tabelle 2 wird jedoch n=29 angegeben. In der 9MIU IFN-a2a sind die Unterschiede sogar noch gravierender: In der Tabelle wird n=28 angegeben, im Text wird jedoch beschrieben, dass von 32 eingeschlossenen (Myhr 1999, S.1050 [25]) fünf Patienten die

Behandlung abbrechen (Myhr 1999, S.1052 [25]), wodurch eigentlich nur 27 Patienten übrigblieben, und falls 19 Patienten ohne Läsionen 65,5% entsprächen, wie in Tabelle 2 angegeben (Myhr 1999, S.1051 [25], Table2), müssten in dieser Gruppe 29 Patienten sein. Derartige Diskrepanzen lassen auf Fehler in der Berichterstattung einer Studie und damit auf mögliche Validitätsprobleme schließen.

Was den Leser ebenfalls skeptisch stimmen kann, sind nicht vollständig nachvollziehbare Einschlusskriterien. Als Beispiel lässt sich hier Optic Neuritis Study Group 1997 [29] (S.1405) nennen: „The criteria for entry into the ONTT included the diagnosis of acute unilateral optic neuritis with visual symptoms of 8 days or less, age between 18 and 46 years, no previous history of optic neuritis or ophthalmoscopic signs of optic atrophy in the affected eye, ...“. Sowohl der Zeitraum acht Tage als auch das Alter bis 46 Jahre scheinen hier relativ willkürlich festgelegt. Normalerweise würde man annehmen, dass übliche Zeiträume wie zum Beispiel eine Woche für das Bestehen der Optikusneuritis oder für die obere Altersgrenze 50 oder auch 60 Jahre verwendet würden. Es besteht also die Möglichkeit, dass die Einschlusskriterien möglicherweise im Nachhinein zu einem Zeitpunkt festgelegt wurden, als die Patienten schon rekrutiert waren. Was in diesem Fall zusätzlich auffällt, ist, dass in einer Zweitpublikation zu der selben Studie als Grenze für das Einschlussalter 45 Jahre oder weniger genannt werden: „... age range, 18 to 45 years inclusive...“ (Optic Neuritis Study Group 1991, S.1673 [29]). Die Kriterien differieren also um ein Jahr. Diese Diskrepanz zusammen mit dem Anschein einer gewissen Willkür könnten die Validität einer Studie einschränken.

Einen weiteren Aspekt, der für Schritt 7 relevant ist, zeigt Patti 1999 [31] (S.286). Dort wird von einer „Table 4“ gesprochen, die aber nicht vorhanden ist, wodurch wertvolle Informationen über Nebenwirkungen fehlen.

Eine MRT-spezifische Einschränkung der Validität findet sich unter anderem in Zhao 2000 [46] (S.201) und Millefiorini 1997 [19] (S.155). Dort wird jeweils beschrieben, dass für die Anfertigung der MR-Bilder verschiedene MR-Geräte verwendet wurden, allerdings ohne zu evaluieren, ob die Ergebnisse dabei vergleichbar waren. So schreibt Zhao 2000 [46] (S.201): „Each center used ... landmarks specific to its scanner...“. Millefiorini 1997 [19] (S.155) erklärt: „MRI studies .. were obtained with different imagers...“. Da in beiden Fällen nicht untersucht wurde, inwiefern sich die Ergebnisse verschiedener MR-Geräte vergleichen lassen, kann der Leser nicht sicher sein, dass die Resultate genauso ausgefallen wären, falls alle Scans mit dem gleichen Typ MR-Gerät durchgeführt worden wären (zu den Möglichkeiten der MR siehe auch Abschnitt 3.2.2. der vorliegenden Arbeit).

Weiterhin kommt es häufiger vor, dass zu wenig ausführliche Angaben zu den angewandten Diagnostikmethoden gemacht werden, bei den eingeschlossenen Studien also vor allem zu der MRT-Anfertigung und -Auswertung. So konnte die Anfertigung von MRI-

Scans unter anderem in Kappos 2000 [95] oder Kastrukoff 1999 [15] nicht nachvollzogen werden. Das ist deshalb ungünstig, da die Reproduzierbarkeit für jemanden, der die Ergebnisse einer Studie klinisch umsetzen möchte, eingeschränkt ist und er sich somit nicht auf die Resultate der verwendeten Studie verlassen kann. Aber auch eine im Allgemeinen geringe Transparenz einer Publikation kann die Validität einer Studie beeinflussen, wie zum Beispiel in Teksam 2000 [40], wo generell die Beschreibung von Patienten, Diagnostik- und Therapiemethoden zu knapp ausfällt oder in Paolillo 1999 [30] (S.444), wo nicht ganz deutlich wird, wann wie viele Patienten welche Diagnostik und Therapie bekamen.

3.1.2.2. Überprüfung der Wichtigkeit

Aufgrund der ermittelten Validität wurden zunächst 18 Studien für eine Berechnung der Kennzahlen ausgewählt. Davon war es in 15 Publikationen (83%) (siehe Tabelle 6 im Anhang) möglich, eine ARR, RRR und NNT auszurechnen, während in 3 Veröffentlichungen (Kastrukoff 1999 [15]; Filippi, J Neurol Sci 2000 [8]; Rohowsky-Kochan 1997 [33]) (17%) die Angaben für eine Berechnung zu knapp waren. Die Ergebnisse sind im Anhang in Tabelle 7 aufgelistet. Dabei enthält die Spalte „Aussage“ die in der jeweiligen Studie auf ihre Wichtigkeit untersuchten Aussagen. Dort finden sich sowohl klinische als auch technische Outcomes, also MRT-Ergebnisse. Wenn eines dieser beiden Zielkriterien hier fehlt, bedeutet das, dass aus den Angaben in der entsprechenden Veröffentlichung für dieses Merkmal keine Kennzahlen zu errechnen waren. In der Spalte „betroffene Patienten“ sind die von dem Zielkriterium betroffenen Studienteilnehmer im Vergleich Placebo (PL)- und Interventionsgruppe (ET) aufgeführt. 40/98 bedeutet zum Beispiel, dass 40 Patienten aus dieser Vergleichsgruppe, die insgesamt 98 Teilnehmer umfasste, das untersuchte Beispiel eines Merkmals aufwiesen. In den dann folgenden Abschnitten der Tabelle finden sich die Angaben für die relative Risikoreduktion (RRR), die absolute Risikoreduktion (ARR), die number-needed-to-treat (NNT) und die dazugehörigen Konfidenzintervalle. Außerdem ist das Ergebnis des CHI-Quadrat-Tests und dessen Signifikanzniveau dargestellt.

Von den untersuchten Studien wurde in elf (65%) eine positive Wirkung der experimentellen Therapie mit zum Teil unterschiedlichen Ausmaßen gefunden. Drei Behandlungen (Anti-alpha4 Integrin-Antikörper; Lenercept; Cladribin) (18%) konnten eine helfende Wirkung nicht beweisen und schaden eher, was sich in einer negativen NNT-Zahl, auch NNH (number-needed-to-harm) genannt, niederschlägt. Genauere Ausführungen zu den verschiedenen Therapiemöglichkeiten bei MS finden sich in einem kurzen Review der Behandlungsoptionen im Abschnitt 3.2 dieser Arbeit.

3.2. Review über Therapiestudien bei Multipler Sklerose

Um die Ergebnisse der in diese Arbeit eingeschlossenen Studien zu verwerthen, werden diese im Folgenden in Form eines kurzen Reviews im Vergleich mit einem bereits in der Literatur vorhandenen systematischen Review (Clegg 2000 [67]) dargestellt. Bei den Wichtigkeitskennzahlen der Zielkriterien (siehe Abschnitt 3.1.2.2.) existieren zwei große Gruppen: Die eine befasst sich mit klinischen Outcomes, die andere mit den MRT-Resultaten. Leider enthielten nur zwei der in das Review eingeschlossenen Arbeiten (Jacobs 2000 [13]; The Once Weekly Interferon for MS Study Group 1999 [42]) ausreichende Angaben, um zu beiden Formen der Zielkriterien NNTs berechnen zu können. Wo immer aus den Angaben der in das Review eingeschlossenen Studien möglich, werden Konfidenzintervalle (CI) und p-Werte angegeben.

3.2.1. Therapieoptionen

Mittlerweile existieren verschiedene Therapieoptionen für die Langzeitbehandlung der MS. In Deutschland zugelassen sind für diesen Zweck derzeit drei verschiedene Interferone (www.dmsg.de [167]), Azathioprin (Gold 2000, S. 76 [84]; www.muskl.de [174]) und das Glatirameracetat (Copaxone®) (www.dmsg.de [167]; www.simmformation.de [177]), früher Copolymer-1 genannt. Weitere Therapien, wie zum Beispiel Immunglobuline, können bei Versagen der herkömmlichen Therapie eingesetzt werden. In den folgenden Abschnitten werden die NNTs der unterschiedlichen zugelassenen und der noch im experimentellen Stadium befindlichen Behandlungsoptionen aus den Studien der vorliegenden Arbeit und aus Clegg 2000 [67] gegenübergestellt und Schlussfolgerungen gezogen.

3.2.1.1. Interferone

Die drei derzeit in Deutschland zugelassenen Interferone sind (www.dmsg.de [167]):

- Die subkutane Gabe von IFN β -1b (Betaferon®/Betaseron®), 8 Mio. Einheiten jeden zweiten Tag;
- Die intramuskuläre Gabe von IFN β -1a in der Dosierung von 6 Mio. Einheiten einmal pro Woche (Avonex®);
- Die subkutane Gabe von IFN β -1a (Rebif®) in der Dosierung von 3 x 6 Mio. Einheiten pro Woche.

Sieben der für die vorliegende Arbeit bewerteten Publikationen erfüllten die Kriterien für einen Einschluss in die Auswertung der Wirkungen von Interferon β , davon vier Veröffentlichungen, in denen MRT-NNTs, aber keine klinischen Kennzahlen errechnet werden konnten (Zhao 2000 [46]; Simon 1998 [36]; Miller 1999 [20]; Li 1999 [18]), eine RCT, in der rein klinische Outcomes zur Verfügung standen (Patti 1999 [31]) und zwei Publikationen, aus deren Angaben beide Formen der NNT-Berechnung möglich waren (The Once Weekly Interferon for MS Study Group 1999 [42]; Jacobs 2000 [13]). Die einzige vorliegende Studie zu Interferon a-2a (Myhr 1999 [25]) konnte in der Validitätsbewertung nur Stufe 1 erreichen und wurde deshalb aus dem Review ausgeschlossen. Auch Clegg 2000 [67] und die ebenfalls Therapieoptionen bei MS zusammenfassenden Werke von Gold 2000 [84] und Wötzel 2000 [158] enthalten keine Evidence zur Wirksamkeit von IFN-a2a.

Eine Auflistung aller Interferon-Studien, aus deren Resultaten Kennzahlberechnungen möglich waren, findet sich in Tabelle 15 und eine Zusammenstellung der Ergebnisse der NNTs in den Tabellen 12, 13, und 14 wobei ausdrücklich auf die Darstellung weiterer Parameter ohne die Möglichkeit einer Berechnung von NNTs verzichtet wird, um den Umfang dieses Reviews in Grenzen zu halten.

Von den für das Review eingeschlossenen Studien der vorliegenden Arbeit untersuchten zwei RCTs Avonex[®] (Interferon β -1a) (Simon 1998 [36]; Jacobs 2000 [13]), zwei befassten sich mit Rebif[®] (ebenfalls Interferon β -1a) (The Once Weekly Interferon for MS Study Group 1999 [42]; Li 1999 [18]), zwei überprüften den Therapieerfolg von Betaseron (Interferon β -1b) (Zhao 2000 [46]; Miller 1999 [20]), und eine evaluierte natürliches Interferon β (nIFN β) (Patti 1999 [31]).

Die Ergebnisse der klinischen NNT-Berechnungen für Patienten mit schubweiser MS (=RRMS) werden in der folgenden, auf der folgenden Seite fortgesetzten, Tabelle denjenigen von Clegg 2000 [67] gegenübergestellt (für genauere Angaben zu NNTs der vorliegenden Arbeit siehe Tabelle 7):

	Studie	NNT berechnet aus	Ergebnis: NNT (wahrer Wert) =
Berechnungen für die vorliegende Arbeit	The Once Weekly Interferon for MS Study Group 1999 [42]; IFN-β1a (Rebif[®]) 22μg oder 44μg je 1x/Woche; RRMS	• Anzahl der Patienten mit moderaten oder schweren Rückfällen in der LD [*] -Gruppe nach 2 Jahren Therapie	19 (18,8)
		• Anzahl der Patienten mit moderaten oder schweren Rückfällen in der HD ^{**} -Gruppe nach 2 Jahren Therapie	12 (11,7)
	Jacobs, NEJM 2000 [13]; IFN-β1b (Betaseron[®]) 30μg 1x/Woche i.m.; beginnende MS	• Anzahl der Patienten mit Entwicklung klinisch definitiver MS nach mindestens 22 Monaten Therapie	7 (6,0)

	Patti 1999 [31]; nIFN-β 6MIU 3x/Woche i.m.; RRMS und SPMS	• Anzahl der RRMS-Patienten mit Rückfällen über zwei Jahre Therapie	5 (4,8)
		• Anzahl der RRMS-Patienten mit einer Verschlechterung oder Gleichbleiben des EDSS-Scores	3 (2,1)
Clegg 2000 [67], S.10-14, S.51-56	Ebers 1998 [194]; IFN-β1a (Rebif®) 22µg oder 44µg 3x/Woche s.c.; RRMS	• Anteil der RRMS-Patienten ohne Rückfall über zwei Jahre Therapie	9 (CI: 5-36)
		Jacobs 1996 [92]; IFN-β1a (Avonex®) 30µg 1x/Woche i.m.; RRMS	• Anteil der RRMS-Patienten ohne Rückfall über zwei Jahre Therapie
	The IFNB Multiple Sclerosis Study Group 1993 [150] und 1995 [149]; IFN-β1b (Betaseron®) 8MIU oder 1,6MIU je-den 2.Tag s.c.; RRMS	• Anteil der RRMS-Patienten ohne Fortschreiten der Erkrankung über 2 Jahre Therapie	8 (CI: 4-34)
		• Anteil der RRMS-Patienten ohne Rückfall über zwei Jahre Therapie	7 (CI: 4-23)
		• Anteil der RRMS-Patienten ohne Fortschreiten der Erkrankung über 2 Jahre Therapie	12 (CI: 5-8)

Tabelle 12: Gegenüberstellung der klinischen NNT-Ergebnisse für IFN-β aus Berechnungen für die vorliegende Arbeit und aus Clegg 2000 [67] für RRMS-Patienten.

In Tabelle 13 werden die Ergebnisse der klinischen NNT-Berechnungen für Patienten mit sekundär progressiver Multipler Sklerose (=SPMS) denjenigen von Clegg 2000 [67] gegenübergestellt (für genauere Angaben zu NNTs der vorliegenden Arbeit siehe Tabelle 7):

	Studie	NNT berechnet aus	Ergebnis: NNT (wahrer Wert) =
Berechnungen für die vorliegende Arbeit	Patti 1999 [31]; nIFN-β 6MIU 3x/Woche i.m.; RRMS und SPMS	• Anzahl der SPMS-Patienten mit einer Verschlechterung oder Gleichbleiben des EDSS-Scores	3 (2,9)
Clegg 2000 [67], S.14-16, S.57-58	Polman 1995 [123] und Kappos 1998 [95]; IFN-β1b (Betaseron®) 8MIU oder 1,6MIU je-den 2.Tag s.c.; SPMS	• Anteil der Patienten ohne Fortschreiten der Erkrankung über 2,5 Jahre Therapie	9 (CI: 6-28)
		• Gesamtwahrscheinlichkeit, nicht von einem Rollstuhl abhängig zu sein nach 2,5 Jahren Therapie	13 (CI: 7 – 50)
		• Anteil der Patienten ohne Rückfälle über 2,5 Jahre Therapie	11 (CI: 6-66)

Tabelle 13: Gegenüberstellung der klinischen NNT-Ergebnisse für IFN-β aus Berechnungen für die vorliegende Arbeit und aus Clegg 2000 [67] für SPMS-Patienten.

Da die in Clegg 2000 [67] eingeschlossenen Studien keine MRT-Resultate als Zielkriterien enthielten, werden in der folgenden Tabelle 14 nur die Ergebnisse der MRT-NNT-Berechnungen der vorliegenden Arbeit aufgelistet (für genauere Angaben zu den NNTs siehe Tabelle 7):

	Studie	NNT berechnet aus	Ergebnis: NNT (wahrer Wert) =
Berechnungen für die vorliegende Arbeit	Simon 1998 [36]; IFN β-1a (Avonex[®]) 30μg 1x/Woche i.m.; RMS	<ul style="list-style-type: none"> Anzahl der RMS-Patienten mit Gd-positiven T1-Läsionen nach 1 Jahr Therapie 	19 (18,7)
	Li 1999 [18]; IFN-β1a (Rebif[®]) 22μg oder 44μg 3x/Woche s.c.; RRMS	<ul style="list-style-type: none"> Anteil der RRMS-Patienten mit T₂-Aktivität über die Studiendauer in der LD^{*)}-Gruppe 	10 (9,3)
		<ul style="list-style-type: none"> Anteil der RRMS-Patienten mit T₂-Aktivität über die Studiendauer in der HD^{**)}-Gruppe 	5 (4,4)
		<ul style="list-style-type: none"> Anteil der RRMS-Patienten mit CU-Läsionen-Aktivität in Kohorte 2 in der LD^{*)}-Gruppe 	5 (4,8)
		<ul style="list-style-type: none"> Anteil der RRMS-Patienten mit CU-Läsionen-Aktivität in Kohorte 2 in der HD^{**)}-Gruppe 	4 (3,7)
	The Once Weekly Interferon for MS Study Group 1999 [42]; IFN-β1a (Rebif[®]) 22μg oder 44μg je 1x/Woche; RRMS	<ul style="list-style-type: none"> Anteil der aktiven MR-Scans in der LD-Gruppe nach 2 Jahren Therapie 	21^{**) (20,7)}
		<ul style="list-style-type: none"> Anteil der aktiven MR-Scans in der HD-Gruppe nach 2 Jahren Therapie 	6^{**) (6,0)}
	Zhao 2000 [46]; IFN-β1b (Betaseron[®]) 8MIU oder 1,6MIU je- den 2.Tag s.c.; RRMS	<ul style="list-style-type: none"> Anzahl der RRMS-Patienten mit annual accumulation of lesion activity (AAA) in der LD-Gruppe nach drei bis fünf Jahren^{*)} 	8 (7,53)
		<ul style="list-style-type: none"> Anzahl der RRMS-Patienten mit annual accumulation of lesion activity (AAA) in der HD-Gruppe nach drei bis fünf Jahren^{*)} 	6 (5,2)
	Miller 1999 [20]; IFN-β1b (Betaseron[®]); SPMS	<ul style="list-style-type: none"> Anzahl der SPMS-Patienten mit aktiven Scans nach 3 Jahren^{*)} 	6 (5,1)
Jacobs, NEJM 2000 [13]; IFN-β1b (Betaseron[®]) 30μg 1x/Woche i.m.;; beginnende MS	<ul style="list-style-type: none"> Anzahl der Patienten mit beginnender MS, die nach 18 Monaten Therapie Läsionen im T1-gewichteten MRT aufwiesen 	4 (3,2)	

^{*)} Angaben über die zeitliche Auswertung nicht eindeutig;

^{**) NNT nur mit Einschränkungen gültig (siehe Text).}

Tabelle 14: MRT-NNT-Ergebnisse für IFN- β aus Berechnungen für die vorliegende Arbeit.

Generell gilt sowohl für die Studien aus der vorliegenden Arbeit als auch für diejenigen aus Clegg 2000 [67], dass ein direkter Vergleich zwischen den einzelnen Publikationen meist nur schwer möglich ist. Dafür gibt es vielfältige Gründe, als Beispiel seien nur die folgenden vier genannt (Clegg 2000, S.IVff. [67]):

1. Die Endpunkte der Studien sind oftmals unterschiedlich
2. Patientencharakteristika wie Schwere der Erkrankung oder Dauer der MS gleichen sich nicht
3. Medikamentendosierungen- und gabe erfolgen auf verschiedene Weise
4. Die Behandlungszeiträume unterscheiden sich.

Hinzu kommt, dass die meisten Studien methodologische Probleme aufweisen. Zweifel an der Verblindung lassen zum Beispiel alle in Clegg 2000 [67] eingeschlossenen IFN-Publikationen aufkommen (Clegg 2000, S.12f. [67]). In Polman 1995 [123], Kappos 1998 [199], The IFNB Multiple Sclerosis Study Group 1993 [150] und 1995 [149] sowie in Jacobs 1996 [92] wurde außerdem die ITT-Analyse vernachlässigt und wurden nicht alle Patienten in die Endauswertung eingeschlossen. Die Auswertungsergebnisse der für die vorliegende Arbeit untersuchten Studien nach dem objektivierten Bewertungsschema sind in folgender Tabelle wiedergegeben (Auszug aus Tabelle 5 im Anhang):

Studie	Randomisierung			Be- rücks. (6)	ITT		Verblindung		
	Rand. (3)	Zeit- punkt (2)	Geh.- haltg. (2)		An- geg. (1)	Durchgef. (4)	Pat. (2)	Ther. (2)	Diagn. (2)
The Once Weekly Interferon for MS Study Group 1999 [42];	3	2	2	6	1	4	2	0	2
Zhao 2000 [46]	3	0	1	5	1	3	1	1	2
Simon 1998 [36]	3	2	2	2	1	4	2	0	2
Miller 1999 [20]	3	2	2	6	1	4	2	1	2
Li 1999 [18]	3	2	2	6	1	3	2	1	2
Jacobs, NEJM 2000 [13]	3	2	1	3	1	3	2	2	2
Patti 1999 [31]	3	2	1	6	0	3	2	0	2

Tabelle 15 (fortgesetzt auf der nächsten Seite): Evaluationsergebnisse nach dem modifizierten Bewertungsschema der in das Review eingeschlossenen IFN-Studien (Auszug aus Tabelle 5). In Klammern die jeweils möglichen Maximalpunktzahlen.

Studie	Gleichbeh. (5)	Ähnlichkeit		Gesamtpunkte (35)	Abwertung auf	wegen	Endergebnis
		Ähnlkt. (5)	Tab. (1)				
The Once Weekly Interferon for MS Study Group 1999 [42]	3	3	1	29	n	n	3
Zhao 2000 [46]	3	3	1	24	n	n	2
Simon 1998 [36]	3	3	1	25	2	teilw. Berücks. d. Pat.	2
Miller 1999 [20]	4	2	1	30	n	n	3
Li 1999 [18]	3	3	1	29	n	n	3
Jacobs, NEJM 2000 [13]	3	5	1	28	n	n	3
Patti 1999 [31]	3	5	1	28	n	n	3
n = keine Abwertung							

Tabelle 15 (fortgesetzt von der vorherigen Seite): Evaluationsergebnisse nach dem modifizierten Bewertungsschema der in das Review eingeschlossenen IFN-Studien (Auszug aus Tabelle 5). In Klammern die jeweils möglichen Maximalpunktzahlen.

Zhao 2000 [46] hatte zum Beispiel mit dem Zeitpunkt der Randomisierung Schwierigkeiten. So wurden 372 Patienten zunächst randomisiert, von denen *anschließend* 342 Studienteilnehmer für die Aufnahme in die Studie anhand von Einschlusskriterien ausgewählt wurden. Jacobs 2000 [13], schloss nur 80,9% aller randomisierten Patienten in die Auswertung ein und erfüllte damit EBM-Schritt 2 nur noch teilweise. Allerdings ist diese Publikation die einzige der sieben RCTs, in der die therapeutisch tätigen Ärzte zumindest wahrscheinlich verblindet waren. Miller 1999 [20] wiederum enthält nur wenige Angaben, um die Ähnlichkeit der Vergleichsgruppen exakt erkennen zu können.

Alles in allem liegt jedoch die methodologische Qualität der hier verwendeten IFN-Studien über dem Durchschnitt aller in die Bewertung eingeschlossenen RCTs. Daher sollte den Ergebnissen trotz mancher Zweifel durchaus Beachtung geschenkt werden.

Insgesamt stehen für die klinische Kennzahlenberechnung drei Studien zu IFN- β 1a (davon zwei zu Rebif[®], eine zu Avonex[®]) bei RRMS zur Verfügung (siehe Tabelle 13).

Die dabei gefundene NNT zur Verhinderung eines Fortschreitens der Erkrankung und von Rückfällen nach zwei Jahren Therapie lag für Avonex bei jeweils acht Patienten. Das bedeutet, dass acht RRMS-Patienten zwei Jahre lang mit Avonex in der untersuchten Dosierung von 6 MIU einmal pro Woche behandelt werden müssen, um bei einem das Auftreten von Rückfällen bzw. das Fortschreiten der Erkrankung zu verhindern.

Für Rebif liegt die NNT zur Verhinderung von moderaten oder schweren Rückfällen bei 19 Patienten in der 22µg-Gruppe und bei 12 in der 44µg-Gruppe. Der Effekt scheint also dosisabhängig zu sein. Ein derartiger Effekt zeigt sich auch in The Once Weekly Interferon for MS Study Group 1999 [42], wo die Reduktion der Rückfallrate durch Rebif und Avonex in verschiedenen Dosierungen verglichen wird (The Once Weekly Interferon for MS Study Group 1999, S.685 [42], Figure 3). Aus Ebers 1998 [194] ergibt sich für die NNT, um bei einem Patienten einen Rückfall durch zwei Jahre Therapie mit Rebif 6 MIU zu verhindern, die Zahl 9 (Clegg 2000 [67], Ss.11,13).

Zu IFN-β1b sind für die Berechnung klinischer NNTs bei RRMS-Patienten zwei Studien geeignet (siehe Tabelle 12).

Laut den Daten aus Jacobs, NEJM 2000 [13], müssen sieben Patienten mit einem ersten demyelinisierenden Ereignis mindestens 22 Monate behandelt werden, um bei einem das Auftreten klinisch definitiver MS zu verhindern. Hier zeigt sich deutlich eine Besonderheit der Rundung bei NNTs. So liegt die errechnete NNT mit einem Wert von 6.037 nur knapp über 6,0, muss aber auf 7 aufgerundet werden, da die Vorschrift besagt, dass NNTs den Wert der nächsthöheren ganzen Zahl annehmen müssen (Sackett 1999, S.109 [134]).

In der zweiten Publikation zu IFN-β1b werden Patienten mit bereits bestehender MS untersucht. The IFNB Multiple Sclerosis Study Group gibt in ihren Publikationen von 1993 [150] und 1995 [149] Werte an, aus denen errechnet werden kann, dass sieben RRMS-Patienten zwei Jahre lang mit 8 MIU pro Woche therapiert werden müssen, um bei einem zusätzlich das Auftreten von Rückfällen zu verhindern. Damit das Fortschreiten der Erkrankung bei einem zusätzlichen Patienten aufgehalten werden kann, müssen 12 Patienten behandelt werden (Clegg 2000, S.12f. [67]). Allerdings äußern Clegg und Kollegen Zweifel an der Validität der Studie zu IFN-β1b, denn „this trial had a number of important methodological limitations“ (Clegg 2000, S.13 [67]).

Natürliches Interferon β ist die dritte untersuchte IFN-Form bei RRMS (siehe Tabelle 12), die jedoch mit Rebif und Avonex praktisch übereinstimmt. „Interferon-b1a ist mit dem humanen IFN-b in der Aminosäuresequenz identisch und ebenso wie das natürliche Interferon mit einem Zuckerrest versehen (glykosiliert). Glykosilierte Interferone (IFN-b1a) sind offenbar besser verträglich als Interferone (IFN-b1b), an die kein Zuckerrest gebunden ist“ (www.ruhr-unibochum.de [176]). Als Ergebnis für die NNT, um bei einem Studienteilnehmer zusätzlich das Auftreten von Rückfällen über zwei Jahre Therapie zu verhindern, ergibt sich die Zahl von fünf Patienten. Noch besser ist der Wert sogar für die NNT zur Verbesserung des EDSS-Scores: Nur drei Patienten müssen hier zwei Jahre lang behandelt werden, um bei einem zusätzlich den EDSS-Score zu verbessern.

Für die SPMS-Behandlung wurden aus zwei Studien klinische NNT-Werte errechnet (siehe Tabelle 13).

Patti 1999 [31] enthielt Daten, aus denen eine NNT von drei Patienten für eine EDSS-Verbesserung nach einer Behandlungsdauer von zwei Jahren mit dreimal wöchentlich 6 MIU natürlichem Interferon β errechnet werden konnte.

Polman 1995 [122] und Kappos 1998 [199] setzten andere Zielkriterien. So beträgt die NNT neun Patienten zur Verhinderung des Fortschreitens der Erkrankung nach 2,5 Jahren Therapie mit 8MIU jeden zweiten Tag. Die NNT, um über den gleichen Zeitraum nicht an den Rollstuhl gefesselt zu werden, liegt bei 13 Patienten, und für die Verhinderung von Rückfällen in dieser Zeit müssen elf SPMS-Patienten therapiert werden (Clegg 2000, S.15 [67]).

Schließlich ist es wichtig, auch auf Studien einzugehen, aus denen NNTs für MR-Parameter errechnet werden konnten. Clegg 2000 [67] enthielt dazu jedoch keine Angaben, sodass die gefundenen Werte ausschließlich auf den in die vorliegende Arbeit eingebundenen Studien beruhen (siehe Tabelle 15). Drei Publikationen enthielten MRT-Daten zu IFN- β 1a, davon Simon 1998 [36] zu Avonex, und Li 1999 [18] sowie The Once Weekly Interferon for MS Study Group 1999 [42] zu Rebif. Weitere drei Veröffentlichungen werteten IFN- β 1b in der Auswirkung auf MRT-Veränderungen aus (Zhao 2000 [46]; Miller 1999 [20]; Jacobs 2000 [13]).

Aus Simon 1998 [36] konnte die NNT für die Verhinderung des Auftretens von Läsionen bei einem zusätzlichen Patienten nach zwei Jahren Therapie mit Avonex 30 μ g (6MIU) i.m. einmal pro Woche mit 19 bestimmt werden.

Aus dem Anteil der Patienten mit T₂-Aktivität im MRI bei Li 1999 [18] wurden für die NNT zur Verhinderung des Auftretens von T₂-Aktivität bei einem zusätzlichen Patienten nach zwei Jahren Therapie mit Rebif 22 μ g (6MIU) s.c. zehn Patienten errechnet und nach Therapie mit 44 μ g fünf Patienten. Außerdem wurde bei Li 1999 [18] in einer Subgruppe mit monatlichen MR-Scans die Anzahl an „combined unique (CU) active lesions“ (Li 1999, S.199 [18]) bestimmt. Dabei handelt es sich um aktive Läsionen, die sowohl im Gd-T₁-gewichteten Bild als auch im PD/T₂-Bild sichtbar sind. Um bei einem Patienten zusätzlich das Auftreten von CU-Läsionen zu verhindern mussten in der 22 μ g-Gruppe fünf und in der 44 μ g-Gruppe vier Patienten zwei Jahre lang behandelt werden. Auch bei The Once Weekly Interferon for MS Study Group 1999 [151] konnten NNTs errechnet werden. Allerdings ist die Aussagekraft in diesem Fall eingeschränkt, da die Datenlage unklar ist. So wird nicht von Patienten mit aktiven Läsionen, sondern von aktiven Scans gesprochen. Um genau zu sein, sagt daher die NNT weniger über die Anzahl an Patienten aus, die zwei Jahre lang behandelt werden müssen, sondern eher über die Anzahl an Scans über zwei Jahre Therapie, um bei einem Scan zusätzlich keine Aktivität auftreten zu lassen. Die NNT von 21 in der 22 μ g-Gruppe und von 6 in der 44 μ g-Gruppe ist daher nur mit Einschränkungen gültig.

IFN- β 1b ist das dritte Interferon. Dazu sind Angaben, um NNTs berechnen zu können, für Patienten mit RRMS, SPMS und beginnender MS in je einer Studie vorhanden. Zhao 2000 [46] gibt den Anteil der Patienten mit „annual accumulation of lesion activity (AAA)“ (Zhao 2000, S.201 [46]) nach drei bis fünf Jahren Therapie mit IFN- β 1b 1,6 MIU s.c. mit 71% und in der HD-Gruppe (8 MIU) mit 66% im Gegensatz zu 84% in der Placebo-Gruppe an (Zhao 2000, S.203 [46]). Daraus lassen sich NNTs von acht Patienten in der LD-Gruppe und sechs Patienten in der Hochdosisgruppe errechnen. Für SPMS-Patienten in Miller 1999 [20] beträgt die NNT nach drei Jahren Therapie mit IFN- β 1b, jeden zweiten Tag 8MIU, sechs Patienten, um bei einem zusätzlich das Auftreten von aktiven Scans während der Behandlungsdauer zu verhindern. Bei Patienten mit beginnender MS konnte nach anfänglich 18-tägiger Prednisontherapie in unterschiedlichen Dosierungen und anschließender 18-monatiger Therapie mit IFN- β 1b in der Dosierung 30 μ g einmal pro Woche i.m. bei einem Patienten zusätzlich das Auftreten von aktiven Scans durch die Behandlung von vier Patienten verhindert werden.

Das Nebenwirkungsprofil der Interferone ist vielschichtig. So kommt es bei allen in bis zu 54% aller Patienten zu Influenza-artigen Symptomen und Kopfschmerzen. Hautreaktionen an der Injektionsstelle sind ebenfalls häufig. Weitere unerwünschte Wirkungen sind neben Ausschlägen Schmerzen und Nekrosen an der Injektionsstelle, Lymphopenie und Leukopenie, Hypertonus, disseminierte Hautausschläge, Myalgien und Muskelhypertonus, Arthralgien, Fieber, Schüttelfrost, Asthenie, trockener Mund, abnorme Leberwerte, Cholelithiasis, Depressionen und Suizidalität (The Once Weekly Interferon for MS Study Group 1999, S.682 [42]; Jacobs 1996, Tabelle 7, S.291 [92] (=Primärpublikation zu Simon 1998); European Study Group on Interferon β -1b in Secondary Progressive MS 1998, S.1495 [72] (=Primärpublikation zu Miller 1999); PRISMS 1998, Tabelle 4, S.1502 [128] (=Primärpublikation zu Li 1999); Jacobs 2000, S.900 [13]; Patti 1999, S.286 [31]).

3.2.1.2. Cladribin

Für die Berechnung von NNT-Werten für Cladribin konnte aus den Studien der vorliegenden Arbeit eine verwendet werden (Rice 2000 [32]). Da Clegg 2000 [67] für die beiden verwendeten Studien keine NNT-Werte für die Cladribin-Behandlung angibt, erfolgt, soweit möglich, ein direkter Vergleich der klinischen Ergebnisse EDSS-Score, SNRS-Score und Exazerbationsrate. Im Folgenden sind die Resultate aufgelistet.

	Studie	Klinische Zielkriterien	Ergebnis:
Berechnungen für die vorliegende Arbeit	Rice 2000 [32]; Cladribin 2,1 mg/kg und 0,7 mg/kg s.c. ; CPMS	• NNT berechnet aus: Anzahl der CPMS-Patienten, die in der LD-Gruppe im Krankheitsverlauf nach einem Jahr Therapie fortgeschritten sind	NNT (wahrer Wert) = -43 (-42,1)
		• NNT berechnet aus: Anzahl der CPMS-Patienten, die in der HD-Gruppe im Krankheitsverlauf nach einem Jahr Therapie fortgeschritten sind	NNT (wahrer Wert) = -35 (-34,2)
		• EDSS-Entwicklung aller Patienten nach einem Jahr Therapie	keine Unterschiede (keine genaueren Ang.)
		• Durchschnittliche EDSS-Zunahme pro SPMS-Patient nach einem Jahr Therapie	Clad HD: ±0,0 Clad LD: ±0,0 PL: +0,3
		• SNRS-Entwicklung aller Patienten nach einem Jahr Therapie	keine Unterschiede (keine genaueren Ang.)
Clegg 2000 [67], S.16-17 und 59-62	Sipe 1994 [205] und Beutler 1996 [190]; Cladribin 2,8 mg/kg i.v. im 1. Jahr, 1,4mg/kg i.v. im 2.Jahr ; CPMS	• mittlere gepaarte Unterschiede des EDSS-Scores der CPMS-Patienten zwischen Placebo und Cladribin-Behandlung nach einem Jahr Therapie	1,3 (CI: 0,6 – 2,0)
		• mittlere gepaarte Unterschiede des SNRS-Scores der CPMS-Patienten zwischen Placebo und Cladribin-Behandlung nach einem Jahr Therapie	-12,5 (CI: -16,7 – (-8,2))
	Romine 1999 [204]; Cladribin 2,1 mg/kg s.c. ; RRMS	• Durchschnittliche EDSS-Zunahme pro RRMS-Patient über 18 Monate Therapie	Clad: +0,4 PL: +0,3
		• Durchschnittliche SNRS-Zunahme pro RRMS-Patient über 18 Monate Therapie	Clad: +2,2 PL: +4,0
		• Exazerbationsraten der RRMS-Patienten zwischen den Therapiemonaten 7 bis 18	Clad: 0,66/Jahr (CI: 0,37 – 1,05) PL: 1,34/Jahr (CI: 0,90 – 1,93)

Tabelle 16: Auflistung der NNT-Ergebnisse für Cladribin aus Berechnungen für die vorliegende Arbeit und Gegenüberstellung mit den Review-Ergebnissen aus Clegg 2000 [67].

Eine NNT konnte nur aus Rice 2000 errechnet werden. Es zeigt sich, dass Cladribin einen nicht signifikanten leicht negativen Effekt auf die Wahrscheinlichkeit der Krankheitsprogression hat (Rice 2000, S.1148 [32], Figure 1). Die negative NNT, also in diesem Fall eine Number-needed-to-harm (NNH), beträgt für die 2,1mg/kg-Gruppe 35 Patienten und für die 0,7mg/kg-Gruppe 43. Wenn diese Anzahl an Patienten also ein Jahr lang Cladribin in der

angegebenen Dosierung erhalte, würde bei einem zusätzlich ein Fortschreiten der Erkrankung ausgelöst. In der durchschnittlichen EDSS- und SNRS-Score-Entwicklung findet sich jedoch kein Unterschied über die Behandlungsdauer, und bei der Subgruppe der SPMS-Patienten errechnet sich eine durchschnittliche Differenz von 0,3 Punkten auf dem EDSS-Score zum Vorteil der Cladribin-Gruppe.

Laut Clegg 2000 [67] leidet die Reliabilität von Sipe 1994 [205] und Beutler 1996 [190] an methodischen Schwierigkeiten (Clegg 2000 [67], S.16f.). So ist die Effektivität der Verblindung unklar und die Cladribin-Dosis wurde im zweiten Studienjahr halbiert, sodass die Effekte in Jahr 1 und Jahr 2 nicht vergleichbar waren. Außerdem wurden Patientenverluste nicht klar genug beschrieben und keine ITT-Analyse durchgeführt (Clegg 2000, S.16f. [67]). Der mittlere gepaarte Unterschied des EDSS-Scores betrug 1,3 zwischen Placebo- und Cladribin-Gruppe, der mittlere gepaarte Unterschied des SNRS-Scores –12,5. Dies würde bedeuten, dass CPMS-Patienten unter Cladribin-Therapie einen günstigeren Verlauf der Erkrankung erwarten könnten (Clegg 2000, S.17 [67]). Romine und Kollegen fanden zwar eine signifikante Reduktion des kombinierten Messwertes aus Häufigkeit und Schwere der Rückfälle für die Cladribin-Behandlungsgruppe (Clegg 2000, S.17 [67]), und die Rate der Exazerbationen wurde gesenkt. Dennoch sind keine statistisch signifikanten Unterschiede des EDSS-Scores oder des SNRS-Scores zu finden. Außerdem ist Cladribin ein potentiell toxisches Immunsuppressivum (Clegg 2000, S.17 [67]).

Als Nebenwirkungen werden angegeben: Knochenmarkssuppression, Thrombozytopenie, Hepatitis B, Salmonellen, milder Herpes zoster, Infektionen des oberen Respirationstrakts, Muskelschwäche, Purpura, Hautreaktionen an der Injektionsstelle, Hypertonie, Rhinitis, Ataxie, Pharyngitis und andere nicht sicher der Therapie zuzuordnende Begleiterscheinungen (Rice 2000, S.1152 [32], Tabelle 4; Clegg 2000, S. 16 [67], Tabelle 11).

3.2.1.3. Monoklonale Antikörperbehandlung

Unter dem Begriff „monoklonale Antikörpertherapie“ wird bei MS in der Regel die Behandlung mit Immunglobulin G verstanden. Da Antegren aber ebenfalls ein monoklonaler Antikörper ist (Tubridy 1999, S.466 [43]), werden die Resultate der experimentellen Therapie mit Antegren in der folgenden Tabelle 17 mit aufgeführt.

	Studie	Klinische und MR-Zielkriterien	Ergebnis:
Berechnungen für die vorliegende Arbeit	Sørensen 1998 [37]; i.v. IgG 2,0g/kg pro Monat; RRMS und SPMS	• NNT berechnet aus: Anzahl aller Patienten mit mindestens einer Exazerbation über die Studiendauer	NNT (wahrer Wert) = 3 (2,6)
		• Anzahl der Exazerbationen während 6 Monaten Therapie	IgG: 11 PL: 19
		• EDSS-Veränderungen aller Patienten nach 6 Monaten Therapie	IgG: ± 0 PL: +0,3
		• Neue Läsionen aller Patienten im Gd-T ₁ -MRT nach 6 Monaten Therapie	IgG: 1,1 ± 2,0 PL: 2,2 ± 4,2
		• Durchschnittliche Änderung in der T ₂ -Läsionenlast nach 6 Monaten Therapie	IgG: -255 mm² ± 1,085 PL: -185 mm² ± 1,295
	Tubridy 1999 [43]; Antegren 3mg/kg Gesamtdosis; RRMS und SPMS	• NNT berechnet aus: Anzahl der RRMS- und SPMS-Patienten mit Exazerbationen nach den ersten 12 Wochen des FU	NNT (wahrer Wert) = 13 (12,6)
		• NNT berechnet aus: Anzahl der RRMS- und SPMS-Patienten mit Exazerbationen nach den zweiten 12 Wochen des FU	NNT (wahrer Wert) = -4 (-3,6)
		• EDSS-Entwicklung aller Patienten nach 12 Wochen FU im Vergleich zu Studienbeginn	Anteg: -0,06 PL: +0,18
		• EDSS-Entwicklung aller Patienten nach 24 Wochen FU im Vergleich zu Studienbeginn	Anteg: -0,02 PL: +0,02
		• Durchschnittliche Entwicklung neuer T ₂ -Läsionen bei allen Patienten nach 12 Wochen FU	Anteg: 0,8 PL: 1,0
		• Durchschnittliche Entwicklung neuer T ₂ -Läsionen bei allen Patienten nach 24 Wochen FU	Anteg: 2,5 PL: 1,7
Clegg 2000 [67], S.23-25 und 75-78	Fazekas 1997 [75]; i.v. IgG 0,15-2,0 g/kg pro Monat RRMS	• EDSS-Entwicklung aller Patienten nach 2 Jahren Therapie im Vergleich zu Studienbeginn	IgG: -0,23 (CI: -0,43 – 0,03) PL: +0,12 (CI: -0,13 – 0,37)
		• Durchschnittliche jährliche Rückfallrate während 2 Jahren Therapie	IgG: 0,52 (CI: -0,32 – 0,72) PL: 1,26 (CI: 0,75 – 1,77)
	Achiron 1998 [47]; i.v. IgG 0,4g/kg alle 2 Tage nach loading dose von insges. 2g/kg RRMS	• EDSS-Entwicklung aller Patienten nach 2 Jahren Therapie im Vergleich zu Studienbeginn	IgG: -0,3 PL: +0,15
		• Jährliche Exazerbationsrate während 2 Jahre Therapie	IgG: 0,59 PL: 1,61
	Sørensen 1997 [206] und 1998 [37]; i.v. IgG	<i>siehe oben in dieser Tabelle.</i>	<i>siehe oben.</i>

Tabelle 17: Auflistung der NNT-Ergebnisse für eine monoklonale Antikörpertherapie aus Berechnungen für die vorliegende Arbeit und Gegenüberstellung mit den Review-Ergebnissen aus Clegg 2000 [67].

Für Sørensen 1998 [37] und Tubridy 1999 [43] war es möglich, NNT-Werte auszurechnen. So betrug die NNT bei Sørensen 1998 [37] für die Behandlung mit IgG nur drei Patienten, um zusätzlich bei einem Patienten durch die sechsmonatige Therapie mit intravenösem Immunglobulin G (2,0g/kg pro Monat) Exazerbationen zu verhindern. Auch die Anzahl der Exazerbationen war vermindert, jedoch nicht signifikant ($p=0,13$), ebenso wie die EDSS-Änderung mit $\pm 0,0$ in der experimentellen Gruppe nicht eindeutig niedriger lag als mit $+0,3$ in der Placebo-Gruppe. Außerdem erreichte die Veränderung in der T₂-Läsionenlast keine Signifikanz (siehe Tabelle 17). Die signifikant niedrigere Zahl an neuen Gd-T₁-Läsionen ist jedoch eines der Outcomes, die überzeugen können. Bei der Auswertung der Daten muss aber beachtet werden, dass diese Studie von vornherein nicht die Power hatte, um Patienten-zentrierte Outcomes zu untersuchen (Clegg 2000, S.23 [67]).

Fazekas 1997 überprüfte den Effekt von i.v. IgG in einer Dosierung von 0,15 bis 2,0 g/kg pro Monat. Die Ergebnisse der EDSS-Änderung in dieser Studie lassen eine Verbesserung der IgG-behandelten Patienten gegenüber der Placebo-Gruppe erkennen, während Achiron 1998 [47] und Sørensen 1998 [37] hier keine signifikanten Resultate produzieren. Die jährliche Exazerbationen-Rate während der Studiendauer war sowohl bei Fazekas 1997 [75] mit einem p-Wert von 0,0037 als auch bei Achiron 1998 [47] mit einem p-Wert von 0,0006 (Clegg 2000, S.23 [67]) in der Experimentalgruppe signifikant niedriger als in der Placebo-Gruppe. Das relative Risiko wird bei Clegg 2000 [67] mit 0,70 (CI: 0,49 -0,92) für Patienten in Fazekas 1997 [75] und mit 0,72 (CI: 0,54 – 0,97) für Patienten bei Achiron 1998 [47] angegeben. Damit lässt sich ein gewisser Nutzen der IgG-Therapie erkennen.

Das Nebenwirkungsprofil: Häufig treten als akute Symptome Kopfschmerzen, urtikarielle Hautreaktionen und Übelkeit auf. Weitere Nebenwirkungen können schwere Ekzeme, Depressionen, Arthralgien, Parästhesien, pathologisches Lachen oder Weinen und das Auftreten von Hepatitis C sein. Auch Ödeme, Schwindelgefühl und Übelkeit kommen vor (Clegg 2000, S.24 [67]; Sørensen 1998, S.1277f. [37]).

Die Therapie mit Antegren, einem humanisierten monoklonalen Antikörper gegen menschliches $\alpha 4\beta 1$ -Integrin (Tubridy 1999, S.466 [43]), wurde in Tubridy 1999 [43] untersucht. Die NNT, um bei einem Patienten zusätzlich das Auftreten von Exazerbationen zu verhindern, beträgt 13 Patienten für die ersten zwölf Wochen der Therapie. Allerdings schlägt das Ergebnis nach 24 Wochen in das Gegenteil um, und durch Behandlung von 4 Patienten mit Antegren kommt es nach diesem Zeitraum bei einem zusätzlich zu dem Auftreten von Exazerbationen. In der EDSS-Entwicklung zeigt sich zwischen beiden Gruppen kein deutlicher Unterschied. Während für das Auftreten neuer T₂-Läsionen nach zwölf Wochen noch ein signifikanter Unterschied zugunsten der Antegren-Gruppe zu verzeichnen war (Tubridy,

S.471 [43]), kehrt sich innerhalb der zweiten 24 Wochen nach Behandlungsbeginn das Ergebnis ins Gegenteil und erreicht mit durchschnittlich 2,5 neuen T₂-Läsionen pro Patient einen höheren Wert als mit 1,7 in der Placebogruppe.

An Nebenwirkungen durch Antegren waren allein Müdigkeit (p=0,047) und Schlaflosigkeit (p=0,05) signifikant unterschiedlich zu den Placebo-behandelten Patienten.

3.2.1.4. TNF-Neutralisation

Zu einer Neutralisation des Tumornekrosefaktors (TNF) kann man einen löslichen TNF-Rezeptor-Antikörper einsetzen: Lenercept®. Dies ist ein Protein, das mit dem TNF-Rezeptor fusionieren kann und im Tierexperiment vor Entwicklung einer Autoimmunenzephalomyelitis schützt (The Lenercept Multiple Sclerosis Study Group 1999, S.457 [41]). Leider existieren in Clegg 2000 [67] keine Daten zu Lenercept und auch Gold 2000 [84] sowie Wötzel 2000 [158] ist dieses Behandlungskonzept noch nicht berücksichtigt. Daher werden in Tabelle 18 ausschließlich die Daten aus The Lenercept Multiple Sclerosis Study Group 1999 [41] dargestellt.

	Studie	Klinische und MR-Zielkriterien	Ergebnis:
Berechnungen für die vorliegende Arbeit	The Lenercept Multiple Sclerosis Study Group 1999 [41]; Lenercept 10, 50 o. 100mg/ Monat; RRMS	• NNT berechnet aus: Anzahl der Patienten mit mindestens einer Exazerbation nach 24 Wochen Studiendauer in der 100mg-Gruppe im Vgl. mit PL-Patienten	NNT (wahrer Wert) = -4 (-3,1)
		• NNT berechnet aus: Anzahl der Patienten mit mindestens einer Exazerbation nach 48 Wochen Studiendauer in der 100mg-Gruppe im Vgl. mit PL-Patienten	NNT (wahrer Wert) = -4 (-3,5)
		• NNT berechnet aus: Anzahl der Patienten mit verschlechtertem NRS-Score nach 48 Wochen Studiendauer in der 100mg-Gruppe im Vgl. mit PL-Patienten	NNT (wahrer Wert) = -10 (-9,9)
		• Jährliche Exazerbationsrate	Len 100: 1,47 Len 50: 1,64 Len 10: 1,00 PL: 0,98
		• EDSS-Veränderung innerhalb 48 Wochen Therapie	Len 100: +0,17 Len 50: -0,03 Len 10: -0,02 PL: +0,01
		• Durchschnittlich neu aufgetretene aktive Läsionen innerhalb der ersten 24 Wochen Therapie	Len 100: 12,0 Len 50: 15,5 Len 10: 8,2 PL: 8,9

Tabelle 18: Klinische und MRT-Ergebnisse für Lenercept aus Berechnungen für die vorliegende Arbeit.

Die Ergebnisse der Untersuchungen zeigen in allen drei berechneten NNT-Werten negative Vorzeichen, also eine NNH. Diese hat den Wert vier für das Auftreten von Exazerbationen bei einem zusätzlichen Patienten in der 100mg-Gruppe. Bei einer Behandlung von vier Patienten mit Lenercept in einer Dosierung von 100mg pro Monat wird also ein Patient weniger über 24 Wochen frei von Exazerbationen bleiben. Gleiches gilt für Rückfälle nach 48 Wochen – auch hier ist die NNH = 4. Außerdem müssten für eine Verschlechterung des NRS-Scores zehn Patienten über eine Dauer von 48 Wochen mit 100mg Lenercept pro Monat behandelt werden, damit sich bei einem zusätzlich der NRS-Score verschlimmern würde.

Die meisten der Parameter in The Lenercept Multiple Sclerosis Study Group 1999 [41], zu denen keine NNTs errechnet werden konnten, bieten auch kein besseres Bild. So liegt die jährliche Exazerbationsrate in den beiden höheren Lenercept-Dosierungen deutlich über der Placebo-Gruppe. Gleiches gilt für die neu aufgetretenen Läsionen: Auch hier finden sich in der 50mg- und 100mg-Gruppe mehr Läsionen als in der 10mg- oder der Placebo-Gruppe. Trotz der ungünstigen Ergebnisse anderer Parameter wurde für die Entwicklung des EDSS-Scores kein deutlicher Unterschied festgestellt.

Nebenwirkungen (The Lenercept Multiple Sclerosis Study Group 1999, Tabelle 6, S.463 [41]): Kopfschmerzen, Hitzewallungen, Übelkeit, Bauchschmerzen, Dyspnoe.

3.2.1.5. Bestrahlung in Kombination mit Kortikosteroiden

Die Kombinationstherapie aus modifizierter totaler Lymphknotenbestrahlung (TLI) und Kortikosteroiden wurde bisher wie Lenercept in nur sehr wenigen Studien untersucht, von denen Cook 1997 [3] in die Bewertung für die vorliegende Arbeit eingeschlossen war. Die Ergebnisse dieser Studie sind in der folgenden Tabelle zusammengestellt:

	Studie	Klinische und MR-Zielkriterien	Ergebnis:
Berechnungen für die vorliegende Arbeit	Cook 1997 [3]; TLI 11x180 cGY + Steroide 30mg/d PMS	• NNT berechnet aus: Anzahl der Patienten, deren EDSS-Score sich 12 Monate nach der Therapie um 1 Punkt verschlechtert hatte	NNT (wahrer Wert) = 3 (2,9)
		• NNT berechnet aus: Anzahl der Patienten, deren EDSS-Score sich 24 Monate nach Therapie um 1 Punkt verschlechtert hatte	NNT (wahrer Wert) = 2 (2,0)
		• NNT berechnet aus: Anzahl der Patienten mit gleichgebliebenen oder verschlechterten MRI-Ergebnissen	NNT (wahrer Wert) = 4 (3,1)
		• Mittlere Exazerbationsrate pro 100 Patientenmonate	TLI: 2,8 S-TLI: 6,1
		• Patienten mit Verbesserung von durch Gadolinium sichtbaren Läsionen	TLI: 12 S-TLI: 5

Tabelle 19: Klinische und MRT-Ergebnisse für TLI und Steroide aus Berechnungen für die vorliegende Arbeit.

Aus den Resultaten in Tabelle 19 wird ein Behandlungserfolg durch modifizierte LK-Bestrahlung deutlich. Die NNT, um bei einem Patienten zwölf Monate nach Therapie eine Verschlechterung um einen Punkt auf der EDSS-Skala zu verhindern, beträgt nur drei, 24 Monate nach Therapie sogar nur zwei Patienten. Leider war aus Cook 1997 [3] nicht die Exazerbationsrate während der Behandlungszeit oder pro Patient abzulesen. Dennoch bieten die gezeigten Werte der mittleren Exazerbationsrate pro 100 Patientenmonate einen guten Anhaltspunkt dafür, dass die TLI auch in diesem Bereich besser als die Placebo-Bestrahlung (sham TLI = S-TLI) abschneidet. Bei der MRT zeigt sich ebenfalls ein signifikanter Vorteil der TLI gegenüber der S-TLI ($p=0,018$). Zwölf Patienten hatten sich bei den Gadoliniumanfärbenden Läsionen in der TLI-Gruppe verbessert, und in der Placebo-Gruppe konnten nur fünf Patienten eine Abnahme der Läsionenzahl oder -größe verzeichnen. Die NNT, um bei einem Patienten zusätzlich eine Verbesserung der Läsionenzahl zu erreichen, beträgt vier Patienten.

Die bei der Behandlung mit TLI in Kombination mit Kortikosteroiden auftretenden unerwünschten Arzneimittelwirkungen waren in Cook 1997 [3]: Harnwegsinfektionen, Infektionen des Respirationstrakts, orale Candidiasis, Müdigkeit, transiente Amenorrhoe, leichte gastrointestinale Beeinträchtigungen, Diarrhoe, Ösophagitis, axillärer Haarverlust und Spontan- oder Fallfrakturen (Cook 1997, S.178f. [3]).

3.2.1.6. Sulfasalazin

Sulfasalazin wurde von den Publikationen der vorliegenden Arbeit in einer Studie untersucht, jedoch findet sich keine Evidence in Clegg 2000 [67], Gold 2000 [84] oder Wötzel 2000 [158]. Tabelle 20 gibt die Daten zur Sulfasalazin-Therapie wieder, beschränkt sich jedoch auf die Werte aus Noseworthy 1998 [27].

	Studie	Klinische und MR-Zielkriterien	Ergebnis:
Berechnungen für die vorliegende Arbeit	Noseworthy 1998 [27].; Sulfasalazin 500mg/Tag über 1 Woche, anschl. Steigerung bis auf 2000 mg/Tag in der vierten Woche; RRMS und PPMS	• NNT berechnet aus: Anzahl aller Patienten mit mindestens einem Rückfall nach 2 Jahren Therapie	NNT (wahrer Wert) = 8 (7,5)
		• NNT berechnet aus: Anzahl aller Patienten mit mindestens einem Rückfall nach 3 Jahren Therapie	NNT (wahrer Wert) = 19 (18,8)
		• Progressionsrate nach 1 Jahr Therapie	Sulf: 23,3 PL: 25,0
		• Progressionsrate nach 2 Jahren Therapie	Sulf: 38,4 PL: 44,1
		• Progressionsrate nach 3 Jahren Therapie	Sulf: 53,1 PL: 49,6
		• Jährliche Rückfallrate aller Patienten während des ersten Jahres Therapie	Sulf: 1,6 PL: 1,5
		• Jährliche Rückfallrate aller Patienten während der ersten zwei Jahre Therapie	Sulf: 0,9 PL: 1,1
		• Anteil aktiver Scans aller Patienten nach 30 Monaten Therapie*)	Sulf: 40,6% PL: 64,7%

*) Einzelergebnis. Werte für 24 und 36 Monate zeigen keinen Unterschied zwischen den Behandlungsgruppen.
Tabelle 20: Klinische und MRT-Ergebnisse für Sulfasalazin aus Berechnungen für die vorliegende Arbeit

Mit Ausnahme der NNT-Werte für die Freiheit von Rückfällen über die Dauer von zwei (NNT=8) und drei Jahren (NNT=19) finden sich in Noseworthy 1998 [27] kaum signifikant positive Resultate. Die jährliche Rückfallrate aller Patienten während des ersten Jahres der Therapie zeigt keinen signifikanten Unterschied, nach zwei Jahren beträgt die jährliche Rate in der Sulfasalazin-Gruppe 0,9, in der Placebo-Gruppe 1,1 ($p=0,04$) (Noseworthy 1998, S.1349 [27]). Als Einzelresultat muss der Anteil aktiver Scans in den jeweiligen Gruppen nach 30 Monaten Therapie (siehe Tabelle 21) angesehen werden. Weder frühere (mit Ausnahme der 6-Monats-MRT-Untersuchungen) noch spätere Werte zeigen einen signifikanten Unterschied der aktiven Scans beider Behandlungsarme (Noseworthy 1998, S.1348 [27]). Am wichtigsten jedoch sind die Resultate für das primäre Zielkriterium der Progressionsrate. Nach einem und nach zwei Jahren Therapie war die Häufigkeit des Fortschreitens der Erkrankung in der Sulfasalazin-Gruppe etwas geringer als in der Placebo-Gruppe. Nach drei Jahren Therapie kehrten sich die Ergebnisse aber ins Gegenteil, und die Progressionsrate lag mit 53,1% in der Sulfasalazin-Gruppe höher als mit 49,6% in der Vergleichsgruppe.

Zu den Nebenwirkungen der Sulfasalazin-Therapie enthält Noseworthy 1998 [27] keine Angaben.

3.2.1.7. AG284

Goodkin 2000 [10] untersuchte primär die Sicherheit und Verträglichkeit von AG284 bei SPMS-Patienten. Daher hatte die Studie von vornherein nicht die Power, um einen signifikanten Behandlungseffekt festzustellen. Auch zu dieser noch im experimentellen Stadium befindlichen Behandlungsform gab es keine Untersuchungen in Clegg 2000 [67], Gold 2000 [84] oder Wötzel 2000 [158]. In Tabelle 21 sind daher nur die Ergebnisse aus Goodkin 2000 [10] aufgelistet.

	Studie	Klinische und MR-Zielkriterien	Ergebnis:
Berechnungen für die vorliegende Arbeit	Goodkin 2000 [10]; AG284 in Dosierungen zwischen 0,6 und 150,0 µg/kg jeden 2.Tag über 5 Tage; SPMS	• NNT berechnet aus: Anzahl der Patienten mit mind. einer Exazerbation über die FU-Zeit von 84 Tagen	NNT (wahrer Wert) = 8 (7,4)
		• EDSS-Veränderung innerhalb 12 Wochen FU	keine Unterschiede (keine genaueren Ang.)
		• 9HPT-Veränderung innerhalb 12 Wochen FU	keine Unterschiede (keine genaueren Ang.)
		• Unterschiede in Gd-T ₁ -MRT-Läsionen	keine Unterschiede (keine genaueren Ang.)

Tabelle 21: Klinische und MRT-Ergebnisse für AG284 aus Berechnungen für die vorliegende Arbeit.

Die einzige RCT zu AG284 hatte nicht die Power, um Behandlungseffekte erkennen zu können (Goodkin 2000, S.1419 [10]). So gab es weder im EDSS- noch im 9HPT-Score zwischen den Gruppen verschiedene signifikante Veränderungen. Auch in dem Auftreten von Gadolinium-positiven Läsionen wurden keine Unterschiede gefunden. Lediglich die NNT, um bei einem Patienten zusätzlich über die Behandlungsdauer mindestens eine Exazerbation zu verhindern, konnte mit acht Patienten ein positives Ergebnis für AG284 bringen. Aufgrund der geringen Anzahl von nur drei Patienten pro Behandlungsgruppe und kurzer Studiendauer sollte dieses Resultat jedoch mit äußerster Vorsicht betrachtet werden.

Folgende Nebenwirkungen traten während der Beobachtungsdauer von zwölf Wochen auf: Hypertonie, Harnwegsinfektionen, Kopfschmerzen und Schmerzen an der Injektionsstelle. Die unerwünschten Wirkungen waren über alle Behandlungsgruppen gleich verteilt (Goodkin 2000, S.1417 [10]).

3.2.1.8. Mitoxantron

Die beiden in Clegg 2000 [67] eingeschlossenen Studien über Mitoxantron entsprechen den in der vorliegenden Arbeit bewerteten. Da das objektivierte Bewertungssystem nach Hassler komplexer und ausführlicher als der bei Clegg 2000 [67] verwendete Jadad-Bewertungsscore ist (Clegg 2000, S.84 [67]), werden für die Validität der Studien unterschiedliche Ergebnisse erhalten. So konnte nur Millefiorini 1997 [19] die Kriterien für einen Einschluss in das hier vorliegende Review erfüllen, während Edan 1997 [4] mit Validitätsstufe 0 evaluiert wurde und damit nicht berücksichtigt werden kann. Die Ergebnisse aus Millefiorini 1997 [19] sind in Tabelle 22 dargestellt.

	Studie	Klinische und MR-Zielkriterien	Ergebnis:
Berechnungen für die vorliegende Arbeit	Millefiorini 1997 [19]; Mitoxantron 8mg/m² 1x/Monat; RRMS	• NNT berechnet aus: Anzahl der Patienten, deren EDSS-Score sich nach 12 Monaten Therapie um mindestens 1 Punkt verschlechtert hatte	NNT (wahrer Wert) = 6 (5,7)
		• NNT berechnet aus: Anzahl der Patienten, deren EDSS-Score sich nach 24 Monaten Therapie um mindestens 1 Punkt verschlechtert hatte	NNT (wahrer Wert) = 4 (3,3)
		• NNT berechnet aus: Anzahl aller Patienten mit mindestens einer Exazerbation im ersten Jahr	NNT (wahrer Wert) = 3 (2,2)
		• NNT berechnet aus: Anzahl aller Patienten mit mindestens einer Exazerbation im zweiten Jahr	NNT (wahrer Wert) = 3 (2,8)
		• Durchschnittliche Anzahl neuer Läsionen während des ersten Jahres der Therapie	Mito: 2,5 (SF=2,3) PL: 3,6 (SF=4,1) (p=0,6)
		• Durchschnittliche Anzahl neuer Läsionen während des zweiten Jahres der Therapie	Mito: 1,8 (SF=3,0) PL: 3,3 (SF=4,1) (p=0,07)
		• Durchschnittliche Anzahl neuer Läsionen während der gesamten zwei Jahre der Therapie	Mito: 3,5 (SF=3,4) PL: 7,3 (SF=8,0) (p=0,05)
		• Durchschnittliche Anzahl sich vergrößernder Läsionen während der gesamten zwei Jahre der Therapie	Mito: 4,3 (SF=4,7) PL: 4,3 (SF=4,7) (p=1,0)

Tabelle 22: Klinische und MRT-Ergebnisse für Mitoxantron aus Berechnungen für die vorliegende Arbeit.

Millefiorini konnte sowohl für klinische als auch MRT-Outcomes deutliche Verbesserungen durch eine Therapie mit Mitoxantron in der Dosierung einmal pro Monat 8mg/m² nachweisen.

So beträgt die NNT, um bei einem zusätzlichen Patienten eine Verschlechterung des EDSS-Scores um mindestens einen Punkt zu verhindern, innerhalb des ersten Jahres der Therapie bei sechs, nach zwei Jahren der Therapie sogar bei nur vier Patienten (Millefiorini 1997, Tabelle 2, S.155 [19]). Auch die NNT, um bei einem weiteren Patienten das Auftreten von Exazerbationen zu verhindern, liegt sowohl im ersten als auch im zweiten Jahr mit drei Patienten sehr günstig. Keine deutliche Verbesserung ergibt sich für die durchschnittliche Anzahl neuer Läsionen während des ersten Jahres der Therapie ($p=0,6$), und auch im zweiten Jahr kann zwar ein Trend in Richtung Mitoxantron festgestellt werden, doch ist das Ergebnis nicht signifikant ($p=0,07$). Erst, wenn beide Jahre zusammen betrachtet werden, gerät die Studie an die Schwelle zur Signifikanz ($p=0,05$). Die durchschnittliche Anzahl sich vergrößernder Läsionen ist in beiden Gruppen nach zwei Jahren identisch.

An unerwünschten Arzneimittelwirkungen werden Übelkeit/ Erbrechen (in 9 Patienten (18%)), Harnwegsinfekte (3 Pat. (6%)), Kopfschmerzen (3 Pat. (6%)), Infektionen des oberen Respirationstrakts (2 Pat. (4%)) und Diarrhoe (1 Pat. (2%)) genannt (Millefiorini 1997, Tabelle 5, S.157 [19]).

3.2.1.9. Weitere Behandlungsmethoden der MS aus der weiterführenden Sekundärliteratur

Außer den bisher genannten Therapieformen bei Multipler Sklerose existieren noch weitere, die in den in die vorliegende Arbeit eingeschlossenen Studien nicht untersucht wurden. Dazu gehören zum Beispiel Glatirameracetat, das früher Copolymer-1 genannt wurde, und Azathioprin. Diese Medikamente sind, wie in Abschnitt 3.2.1. beschrieben, bereits für die Behandlung der MS in Deutschland neben drei Beta-Interferonen zugelassen.

3.2.1.9.1. Glatirameracetat

Die Ergebnisse für Glatirameracetat aus Clegg 2000 [67] sind in der folgenden Tabelle 23 zusammengefasst.

	Studie	Klinische Zielkriterien	Ergebnis:
Clegg 2000 [67], S.21-23 und 73-74	Nicholson & Milne 1996 (systematisches Review) [203]; Copolymer 1, 20 mg/Tag; RRMS und ERMS	• NNT berechnet aus: Anteil der Patienten mit verbessertem EDSS-Score um mindestens 1 Punkt nach zwei Jahren Therapie, Ergebnisse aus Johnson 1995	Cop1: 24,8% PL: 15,2% NNT = 10
		• NNT berechnet aus: Anteil der Patienten mit Krankheitsprogression innerhalb zwei Jahren Therapie, Ergebnisse aus Bornstein 1987	Cop1: 20,0% PL: 47,8% NNT = 4
		• Anteil der Patienten ohne Fortschreiten der Erkrankung innerhalb zwei Jahren Therapie, Ergebnisse aus Johnson 1995	Cop1: 78,4% PL: 75,4%
		• Jährliche Exazerbationsrate während 2 Jahren Therapie, Ergebnisse aus Johnson 1995	Cop1: 1,19 PL: 1,68 (p=0,007)
		• Jährliche Exazerbationsrate während 2 Jahren Therapie, Ergebnisse aus Bornstein 1987	Cop1: 0,6 PL: 2,7 (p nicht angeg.)
		• Anteil der Patienten ohne Rückfälle innerhalb zwei Jahren Therapie, Ergebnisse aus Johnson 1995	Cop1: 33,6% PL: 27,0% (p=0,098)
		• Anteil der Patienten ohne Rückfälle innerhalb zwei Jahren Therapie, Ergebnisse aus Bornstein 1987	Cop1: 56% PL: 26% (p=0,045)

Tabelle 23: Klinische Ergebnisse für Glatirameracetat aus Clegg 2000 [67].

Das systematische Review zu Glatirameracetat von Nicholson & Milne 1996 [203] enthält laut Clegg 2000 [67] qualitätsbeeinflussende Fehler. So wurde die Suchstrategie nicht erläutert, und es fehlten Einschluss- und Qualitätskriterien. Außerdem wurden durch den zusätzlichen Bericht über neuropsychologische Resultate der Patienten aus Johnson 1995 [197]

(Weinstein 1999 [209]) methodische Schwierigkeiten der ursprünglichen Studie sichtbar: Es fehlte eine Beschreibung der Randomisationsmethoden und der Verblindung (Clegg 2000, S.21 [67]). Trotz eingeschränkter Validität sollen die Ergebnisse hier kurz vorgestellt werden. Die Resultate aus Weinstein 1999 [207] werden, da kein Vergleich mit den Ergebnissen anderer Studien möglich ist, nicht berichtet.

Um bei einem Patienten zusätzlich eine Verbesserung des EDSS-Scores zu erreichen, müssen zehn MS-Patienten zwei Jahre lang mit Glatirameracetat in einer Dosierung von 20 mg/Tag behandelt werden. Nur vier Probanden müssen laut Bornstein 1987 [191] therapiert werden, um bei einem zusätzlich über zwei Jahre eine Krankheitsprogression zu verhindern (Clegg 2000, S. 21 [67]). Laut Johnson 1995 [197] gibt es zwischen experimenteller und Placebo-Gruppe jedoch kaum einen Unterschied, was den Anteil der Patienten ohne Fortschreiten der MS betrifft (Clegg 2000, S. 21 [67]). Auch die jährliche Exazerbationsrate ist in beiden Studien aus Nicholson und Milne 1996 [203] unterschiedlich. Johnson berichtet über 1,19 Exazerbationen pro Jahr in der Glatirameracetat-Gruppe und über 1,68 Rückfälle in der Placebo-Gruppe ($p=0,007$) (Clegg 2000, S.21 [67]). Bornstein stellt, ohne jedoch einen p-Wert dafür anzugeben, einen noch größeren Unterschied fest: 2,7 Exazerbationen pro Jahr für die Placebo-Patienten und nur 0,6 Ereignisse in der experimentellen Gruppe (Clegg 2000, S.21 [67]). Auch für den Anteil der Patienten ohne Rückfälle finden sich Unterschiede, die wiederum größer in Bornstein 1987 [191] ausfallen: 56% in der Glatirameracetat-Gruppe sind über zwei Jahre frei von Rückfällen, bei Johnson 1995 [197] sind es nur 33,6%. Die Placebo-Patienten bleiben bei Bornstein in 26% von Exazerbationen verschont, bei Johnson in 27%. Aufgrund der unterschiedlichen Resultate ist die Differenz zwischen den Behandlungsgruppen für die Exazerbationen nur in Bornstein 1987 [191] signifikant mit $p=0,045$.

Als Nebenwirkungen der Glatirameracetat-Therapie werden genannt:: Lokalisiertes Erythem und Verhärtung an der Injektionsstelle (90% der Patienten), vorübergehende systemische Reaktionen (15%). Diese Symptome treten alle auch in der Placebo-Gruppe auf, jedoch mit einer niedrigeren Inzidenz (56% für lokale, 3% für systemische Reaktionen) (Clegg 2000, S.21f. [67]).

3.2.1.9.2. Azathioprin

Die dritte derzeit in Deutschland für den Einsatz bei MS zugelassene Therapieoption bietet Azathioprin. Die Ergebnisse aus Clegg 2000 [67], S.9f. werden in Tabelle 24 dargestellt.

	Studie	Klinische Zielkriterien	Ergebnis:
Clegg 2000 [67], S.9-11 und 47-50	Milanese 1993 [201]; Azathioprin, 2 mg/kg/Tag; RRMS, PMS und RPMS	• Mittlere Veränderung des EDSS-Scores nach einem Jahr Therapie	Aza: 0,29 PL: 0,42
		• Mittlere Veränderung des EDSS-Scores nach zwei Jahren Therapie	Aza: 0,26 PL: 0,84
		• Mittlere Veränderung des EDSS-Scores nach drei Jahren Therapie	Aza: 0,27 PL: 1,22
		• Relatives Risiko eines Rückfalls nach 3 Jahren Therapie für PL gegenüber Aza	RR=1,6 (CI: 1,07 – 2,49)
	Yudkin 1991 (systematisches Review) [210]; Azathioprin, 2-3 mg/kg/Tag; RRMS, RPMS und PMS	• Unterschiede im EDSS-Score nach 1 Jahr Therapie	Aza-PL: -0,03 (CI: -0,18 – 0,12) p=n.s.
		• Unterschiede im EDSS-Score nach 2 Jahren Therapie	Aza-PL: -0,22 (CI: -0,43 – 0,003) p<0,06
		• Unterschiede im EDSS-Score nach 3 Jahren Therapie	Aza-PL: -0,24 (CI: -0,51 – 0,03) p<0,09
		• Odd's ratio für die Freiheit von Rückfällen innerhalb eines Jahres Therapie	OR=1,51 (CI: 1,12 -2,03) (p<0,01)
		• Odd's ratio für die Freiheit von Rückfällen innerhalb zwei Jahre Therapie	OR=2,04 (CI: 1,42 – 2,93) (p<0,01)
		• Odd's ratio für die Freiheit von Rückfällen innerhalb drei Jahre Therapie	OR=1,97 (CI: 1,27 -3,04) (p<0,01)

Tabelle 24: Klinische Ergebnisse für Azathioprin aus Clegg 2000 [67].

Die klinischen Ergebnisse aus Clegg 2000 [67] beinhalten die Resultate zweier Studien: Bei Yudkin 1991 [210] handelt es sich um ein systematisches Review aus sieben Studien, Milanese 1993 [201] hat die Validitätseinschränkung, dass darin keine ITT durchgeführt wurde. Steck 1990 [207], eine weitere in Clegg 2000 [67] eingeschlossene Publikation führte eine Open-label-Therapie durch und wurde daher aus der vorliegenden Arbeit ausgeschlossen. Nachdem Yudkin 1991 [210] nur äußerst geringe Unterschiede des EDSS-Scores nach einem Jahr Therapie mit Azathioprin in der Dosierung 2 bis 3 mg/kg/Tag feststellt, verbessert sich der Wert nach zwei Jahren auf einen durchschnittlichen Vorteil des EDSS-Scores in der Azathioprin-Gruppe von 0,22 (p<0,06). Nach drei Jahren schließlich ist der Unterschied um 0,02 auf 0,24 angestiegen, der p-Wert hat sich mit 0,09 jedoch wieder etwas verschlechtert.

Ohne Angabe von CI oder p-Wert finden sich zum Vergleich in Milanese 1993 [201] folgende Werte für die durchschnittliche Veränderung des EDSS-Scores: Nach einem Jahr Therapie liegt Azathioprin um 0,13 Bewertungspunkte besser als Placebo, nach zwei Jahren Therapie Azathioprin um 0,58 vor Placebo und nach drei Jahren um 0,95. Alle drei Werte sind jedoch nicht signifikant.

Deutlich stellt sich immerhin die Odd's ratio für die Freiheit von Rückfällen nach ein, zwei und drei Jahren Therapie jeweils mit einem p-Wert von unter 0,01 dar (Clegg 2000, S.9 [67]), und auch das relative Risiko eines Rückfalls nach drei Jahren Therapie liegt für Placebo 1,6-fach höher als für Azathioprin (Clegg 2000, S.9 [67]).

An Nebenwirkungen wurden berichtet: Leukopenie, Anorexie, Diarrhoe, Erbrechen, Bauchschmerzen, gastrointestinale Probleme, erhöhte Leberwerte, Hautausschläge, Herpes zoster, Panzytopenie und Makrozytose (Clegg 2000, S.9f. [67]). Als besonders gravierend ist unerträgliches Erbrechen einzustufen, das bei etwa 11% aller Azathioprin-Patienten auftritt (Clegg 2000, S.10 [67]).

3.2.1.9.3. Cyclophosphamid

Zu Cyclophosphamid wurden in Clegg 2000 [67] fünf Studien eingeschlossen, von denen eine nicht verblindet und zwei nur einfach-verblindet stattgefunden hatten und daher aus der Analyse für die vorliegende Arbeit ausgeschlossen wurden. Die Resultate der zwei übrigen Studien sind in der folgenden Tabelle dargestellt.

	Studie	Klinische Zielkriterien	Ergebnis:
Clegg 2000 [67], S.17-20 und 63-71	Canadian Cooperative Multiple Sclerosis Study Group 1991 [192]; Gruppe 1: Cyclophosphamid 1g i.v bis 9g Gesamtdosis.+ anschl. Prednisolon 20mg oral jeden 2.Tag oder WBC < 4,5 x 10⁹/l Gruppe 2: Plasmaaustausch + orales Cyclophosphamid 1,5 – 2g/kg/Tag .+ Prednisolon 20mg oral jeden 2.Tag Gruppe 3: Placebo Dauer in Grp 2+3 je 22 Wochen; CPMS und RRMS	• mittlere Veränderung des EDSS-Scores nach drei Jahren	Cycl: 0,81 (SF=0,14) Plasmaust: 0,69 (SF=0,11) PL: 0,69 (SF=0,10)
	Killian 1988 [200]; Cyclophosphamid 750 mg/m² 1x/Monat; RRMS	• Unterschiede im EDSS-Score nach 1 Jahr Therapie	keine Unterschiede (keine genaueren Ang.)
		• Unterschiede im Ambulation Index nach 1 Jahr Therapie	keine Unterschiede (keine genaueren Ang.)
		• Durchschnittliche Anzahl an Rückfällen innerhalb 1 Jahr Therapie	Cycl: 0,5 (SF 0,2) PL: 2,3 (SF 0,6) (p=0,06)

Tabelle 25: Klinische Ergebnisse für Cyclophosphamid aus Clegg 2000 [67].

Aufgrund der ungünstigen Studienlage hinsichtlich der Validität der Cyclophosphamid-Studien konnten nur zwei der fünf in Clegg 2000 [67] eingeschlossenen Publikationen für die Auswertung der Behandlungswirkungen von Cyclophosphamid verwendet werden, wobei selbst bei diesen beiden Studien Schwierigkeiten auftreten. Die Canadian Cooperative Multiple Sclerosis Study Group 1991 [192] berichtete zum Beispiel nicht über Studienabbrüche und Killian 1988 [200] schloss nur insgesamt 14 Patienten ein, weswegen die Ergebnisse keine Signifikanz erreichen konnten. Trotzdem sollen die Resultate kurz beschrieben werden. Die Canadian Cooperative Multiple Sclerosis Study Group untersuchte in Gruppe 1

Cyclophosphamid mit anschließender Steroidgabe im Vergleich zu Gruppe 2 mit einem Plasmaaustausch + Cyclophosphamid + Steroide und Gruppe 3 mit Placebogabe. Das einzige Ergebnis der mittleren Veränderung des EDSS-Scores drei Jahre nach Studienbeginn war mit 0,81 für Gruppe 1, 0,69 für Gruppe 2 und 0,69 für Gruppe 3 nicht signifikant unterschiedlich zwischen den Therapiegruppen (Clegg 2000, S.18 und 65f. [67]). Auch Killian konnte weder für EDSS-Score noch Amublation Index Unterschiede feststellen. Lediglich die durchschnittliche Zahl an Rückfällen war in der Experimentalgruppe mit 0,5 für den Zeitraum des Studienjahres um 1,8 niedriger als in der Placebogruppe. Wegen der geringen Zahl an Patienten erreichte jedoch auch dieses Ergebnis keine Signifikanz ($p=0,06$) (Clegg 2000, S.19 und 67f. [67]).

Als Hauptnebenwirkungen werden Erbrechen, Übelkeit, Alopezie, Amenorrhoe und Urtikaria genannt (Clegg 2000, S.20 [67]). Außerdem traten hämorrhagische Zystitis, Sepsis, Diabetes, Herpes zoster, Lungenembolie, Angina und Tod durch akute Bronchopneumonie auf (Clegg 2000, S.18 [67]).

3.2.1.9.4. Methotrexat

Für die Behandlung mit Methotrexat konnten zwei Studien für den Einschluss in das systematische Review von Clegg 2000 [67] identifiziert werden. Da keine der in die vorliegende Arbeit eingeschlossenen Publikationen die Therapiewirkung von Methotrexat untersuchte, werden in Tabelle 26 als Grundlage für eine Beurteilung der Effektivität die Ergebnisse der Studien aus Clegg 2000 [67] dargestellt.

	Studie	Klinische Zielkriterien	Ergebnis:
Clegg 2000 [67], S.25f. und 79-81	Currier 1993 [193]; Methotrexat 1x/Woche: 2,5 mg alle 12 Stunden für 3 Dosen; definitive MS	• Anzahl aller Patienten mit Exazerbationen über 18 Monate Therapie	Meth: 8/22 Pat.=36,3% PL: 9/22 Pat.=40,9%
		• Verschlechterung der Subgruppe der RRMS-Patienten im EDSS-Score nach 18 Monaten Therapie	Meth: 3/9 Pat.=33,3% PL: 5/11 Pat.=45,4%
	Goodkin 1995 [195]; Methotrexat 7,5 mg oral 1x/Woche; CPMS	• Scheitern der Therapie ^{*)}	Meth: 16/31 Pat.=51,6% PL: 24/29 Pat.=82,8% (p=0,011) (OR=4,5 (CI:1,36 – 14,85))
		• Scheitern der Therapie ^{*)} für den EDSS-Score alleine	Meth: 11/31 Pat.=35,5% PL: 15/29 Pat.=51,7% (p=0,205)

^{*)} Scheitern der Therapie definiert als Auftreten eines oder mehrerer der folgenden Möglichkeiten für mehr als zwei Monate: -Verschlechterung im EDSS-Score um mindestens 1,0 Punkte bei Eingangsscore zwischen 3,0 und 5,0 oder um mindestens 0,5 Punkte bei Eingangsscore zwischen 5,5 und 6,5; -Verschlechterung im Ambulation Index-Score von 2 bis 6 um mindestens einen Punkt; -Verschlechterung im box-and block/9-hole peg test um mindestens 20%

Tabelle 26: Klinische Ergebnisse für Methotrexat aus Clegg 2000 [67].

Für die Gesamtzahl aller Patienten mit Exazerbationen kann Currier 1993 [193] keinen Behandlungseffekt feststellen. In der Subgruppe der RRMS-Patienten wird ein leichter Trend mit Vorteil für Methotrexat deutlich (Clegg 2000, S. 25f. [67]). Goodkin 1995 [195] weist mit 51,6% bei Methotrexattherapie und 82,8% in der Placebogruppe für das zusammengesetzte Merkmal des Therapiemisserfolgs (siehe Tabelle 26) einen signifikanten Effekt von Methotrexat im Vergleich zur Placebogruppe auf (p=0,011). Allerdings existiert alleine für eine Verschlechterung des EDSS-Scores als Teil des Misserfolgskonzepts mit p=0,205 keine Signifikanz.

Nebenwirkungen: mittelstarker Haarausfall, abnorme Leberwerte, Infektionen des oberen Respirationstrakts, Harnwegsinfekte, Übelkeit, Kopfschmerzen, Fieber, mukokutaner Herpes, Muskelschmerzen, Rückenschmerzen, Maldigestion und Diarrhoe (Clegg 2000, S.25f. [67]).

3.2.1.9.5. Cyclosporin A

Cyclosporin A als ein weiteres immunmodulatorisches Medikament wurde lediglich in Gold 2000, aber weder in den Studien der vorliegenden Arbeit noch in Clegg 2000 [67] abgehandelt und soll daher nur kurz vorgestellt werden. Laut Gold 2000 [84] wurde in einer Therapie-studie von Kappos und Kollegen (Kappos 1988 [198]) „... ein der Azathioprinbehandlung vergleichbarer Effekt festgestellt“ (Gold 2000, S.79 [84]). Allerdings sind die unerwünschten Wirkungen wesentlich stärker vorhanden, weshalb Cyclosporin A zur Monotherapie bei MS nicht empfohlen wird (Gold 2000, S.79 [84]). Auch in einer amerikanischen Studie (The MS Study Group 1990 [207]) wurde zwar eine verzögerte EDSS-Progression beschrieben, allerdings ebenfalls verbunden mit starken Nebenwirkungen (Gold 2000, S.79 [84]). In Kombina-tionsbehandlungen konnte bisher kein deutlicher Effekt gefunden werden (Gold 2000, S.79 [84]).

Die Nebenwirkungen äußern sich in einem Anstieg des Kreatinins auf mehr als 50% des Ausgangswertes und einer arteriellen Hypertonie. Außerdem kommen Hypertrichose und Gingivahyperplasie vor (Gold 2000, S.79 [84]).

3.2.1.9.6. Glukokortikosteroide

Für die Langzeittherapie mit Glukokortikosteroiden lag in Clegg 2000 [67] keine Beschreibung vor. Die in die vorliegende Arbeit eingeschlossen Studien zu einer Behandlung mit Glukokortikosteroiden (Optic Neuritis Study Group 1997 [29]; Oliveri 1998 [28]; Edan 1997 [4]) erbrachten keine ausreichend valide Evidence, um aussagefähig genug für eine Beurteilung des Therapieerfolgs zu sein. Unter dem Vorbehalt der unter Umständen ebenfalls nicht ausreichenden, aber hier nicht untersuchten Evidence werden deshalb die Ergebnisse aus Gold 2000 [84] kurz dargestellt.

Bei der Verwendung von Kortikosteroiden sollte auf eine möglichst geringe mineralokortikoide Wirkung geachtet werden, um die Nebenwirkungen gering zu halten (Gold 2000, S.63 [84]). Barnes und Kollegen sowie Milligan et al. untersuchten 1985 und 1987 erstmals den Erfolg einer „... intravenösen Hochdosistherapie („Pulstherapie“ mit 500 mg Methylprednisolon über 5 Tage) ...“ (Gold 2000, S.63 [84]) (Barnes 1985 [189]; Milligan 1987 [202]). Dabei konnte ein positiver Effekt sowohl auf rezidivierende als auch auf progrediente MS festgestellt werden. Beck und Kollegen verglichen 1992 die konventionelle Behandlung mit der hochdosierten Pulstherapie und fanden ein schnelleres Abklingen der Initialsymptome bei Patienten mit Optikusneuritis, während die Unterschiede sechs Monate später nur noch bei Gesichtsfeld und Farbsinn deutlich waren (Gold 2000, S.63 [84]; Beck 1992 [61]). Weitere Studien zur Glukokortikoidbehandlung leiden an geringer Validität oder verminderter Aussagekraft (Gold 2000, S.63 [84]).

Derzeit wird Methylprednisolon für die Therapie bei Schüben und versuchsweise bei fortschreitendem Krankheitsverlauf eingesetzt, auch wenn die Patienten eigentlich stabil erscheinen (Gold 2000, S.64 [84]).

„... Die Nebenwirkungen (sind bei Pulstherapie) gering und meist gut beherrschbar“ (Gold 2000, S.64 [84]). Lediglich bei Langzeittherapie können Infektionsneigung, Adipositas, Ödeme, Hypokaliämie, Diabetes mellitus, arterielle Hypertonie, Katarakt, peptisches Ulkus, Osteoporose und psychische Veränderungen auftreten (Gold 2000, S.63 [84]).

3.2.2. Bedeutung der Magnetresonanztomographie bei Multipler Sklerose

Die MRT ist eine der wenigen objektiven Möglichkeiten der Diagnostik bei Multipler Sklerose, weshalb Ergebnisse einer MRT-Untersuchung als primäres oder sekundäres Zielkriterium Voraussetzung für den Einschluss von Studien in die vorliegende Arbeit waren. Allerdings ist die Bedeutung der MRT bei MS immer noch umstritten. Daher wird hier der derzeitige Stand der Wissenschaft über die Möglichkeiten und Limitationen der MRT als Untersuchungsmethode für die MS kurz dargestellt.

3.2.2.1. Korrelation von Magnetresonanztomographie-Ergebnissen mit dem Behinderungsgrad von MS-Patienten

Grundsätzlich stellt sich für die MRT die Frage, ob sie ein geeignetes Mittel zur Überprüfung der Schwere der Krankheit und damit auch geeignet zur Evaluation des Therapieerfolgs in MS-Studien ist. Daher soll hier der Zusammenhang zwischen MRT und EDSS sowie zwischen MRT und Schubrate vorgestellt werden.

Laut Filippi und Kollegen korreliert die T₂-Läsionenlast bei Patienten mit klinisch definitiver MS zu Beginn ihrer Studie mit der Schwere der Erkrankung nach fünf Jahren, dem Zeitpunkt der FU-Untersuchung ($p < 0,0001$) (Filippi 1994, S.637 [77]). In einem Review von neun Studien aus den Jahren 1992 bis 1995 konnten Kappos und Kollegen jedoch keine statistisch relevante Korrelation zwischen der Anzahl von Gd-aufnehmenden Läsionen eines Scans zu Beginn einer Studie und der Entwicklung des EDSS-Scores nach ein oder zwei Jahren Beobachtungszeit feststellen (Kappos 1999, S.966 [95]). Die mittlere Anzahl an Läsionen in monatlichen MR-Aufnahmen während des ersten halben Jahres einer Studie ist immerhin schwach prädiktiv für den EDSS-Zuwachs nach zwölf Monaten ($p = 0,082$) und kann mit einer etwas größeren Wahrscheinlichkeit durch logistische Regression die EDSS-Steigerung nach 2 Jahren voraussagen ($p = 0,049$) (Kappos 1999, S.966 [95]). Dies gilt ebenfalls für alle Subgruppen, wie zum Beispiel RRMS- oder SPMS-Patienten sowie Teilnehmer an Studien zum natürlichen Verlauf der MS und an Therapiestudien. Auch Gruppeneinteilungen nach Alter, Geschlecht oder Dauer der Krankheit veränderten die Ergebnisse nicht (Kappos 1999, S.966 [95]).

Für viele wichtige Zielkriterien konnte keine Korrelation zwischen MRT und klinischer Symptomatik der untersuchten Patienten gefunden werden (Miller 1996, S.8 [104]). Dennoch existieren eine Reihe von Zusammenhängen. So beschreiben Miller und Kollegen zum Beispiel eine Korrelation zwischen klinischem Bild und anatomischer Lokalisation der Läsion bei Untersuchung einer Stelle mit klinischer Relevanz, wie des Nervus opticus oder des Fascicu-

lus longitudinalis (Miller 1996, S.8 [104]). Auch bei neuropsychologischen Ausfällen werden höhere Läsionenlasten gefunden als bei Patienten ohne derartige Pathologie (Miller 1996, S.8 [104]). Dies bestätigen auch Filippi et al. 1995 [78], S.1607. Allerdings besteht nur ein mäßiger Zusammenhang zwischen T₂-Läsionenlast und EDSS-Score in Kohorten mit unterschiedlichen Gruppengrößen und verschiedener Dauer der Erkrankung (Miller 1996, S.8 [104]), während eine bessere Korrelation in einer homogeneren Gruppe von 38 Patienten gefunden wurde (Filippi 1994, S.637f. [77]). Auch Paty und Kollegen fanden in einer Studie mit 327 MS-Patienten eine Korrelation zwischen Gesamtläsionenzahl und EDSS-Score sowohl zu Beginn (R=0,231) als auch am Endpunkt nach drei Jahren (R=0,262) (Paty 1993, S.665 [120]), ebenso wie Gass und Kollegen bei 43 Patienten eine Korrelation beschreiben (p=0,03) (Gass 1994, S.63f. [82]). Laut Miller und Kollegen kommen neue oder Gd-aufnehmende Läsionen während MS-Schüben häufiger vor (Miller 1996, S.9 [104]).

Hinsichtlich der prädiktiven Aussage fanden Kappos und Kollegen, dass die Anzahl an Gd-aufnehmenden Läsionen zu Beginn einer Studie bei Patienten mit bereits ausgeprägter klinisch definitiver MS einen Vorhersagewert für die Rückfallrate ein Jahr später hat (RR pro 5 Läsionen=1,13; p=0,04) (Kappos 1999, S.966 [95]). Dies gilt jedoch nur mit Einschränkung. Denn nach Durchführung einer Regressionsanalyse der Ergebnisse verschiedener Studien als Cofaktor war nur noch die mittlere und die Standardabweichung der Läsionenzahl in monatlichen Scans der ersten sechs Monate einer Studie signifikant korreliert mit der späteren Anzahl an Schüben (Kappos 1999, S.966 [95]).

3.2.2.2. Prädiktiver Wert der Magnetresonanztomographie bei Patienten mit einem ersten demyelinisierenden Ereignis

Miller und Kollegen beschreiben in ihrem Review 1996 eine hohe Korrelation zwischen dem Auftreten von T₂-Läsionen während eines ersten demyelinisierenden Ereignisses und der Entwicklung von Multipler Sklerose während der nächsten ein bis fünf Jahre (Miller 1996, S.7 [104]). Die Häufigkeit klinisch definitiver MS schwankt dabei zwischen 23% (Frederiksen 1991, S.343 [79]) und 93% (Ford 1992, S.250 [79]). Bei Patienten ohne einen pathologischen MR-Befund entwickelt sich hingegen nur in 4% eine definitive MS (Miller 1996, S.7 [104]). O'Riordan, Sailer und Kollegen konnten in einer weiteren Studie 1998 bei 72 Patienten mit Entwicklung einer klinisch definitiven MS nach anfänglich isolierten Symptomen eine hochsignifikante Korrelation zwischen Läsionenzahl zu Beginn und EDSS-Score nach fünf und zehn Jahren Beobachtungszeit feststellen (p<0,001) (O'Riordan 1998, S.499 [117]; Sailer 1999, S.603 [137]). Auch die Anzahl neuer Läsionen über die ersten fünf Jahre korrelierte mit der Änderung des EDSS-Scores in dieser Zeit (O'Riordan 1998, S.499 [117]). Ein

schwächerer aber immer noch signifikanter Zusammenhang zwischen neuen Läsionen und EDSS bestand im Zeitraum fünf bis zehn Jahre nach einem ersten demyelinisierenden Ereignis ($p=0,03$) (O'Riordan 1998, S.499 [117]). Mit wachsender Anzahl und Größe der Läsionen wächst auch das Risiko für ein späteres Entstehen einer klinisch definitiven MS (Miller 1996, S.7 [104]). Auch eine infratentorielle Lage der Läsionen birgt eine hohe Progressionswahrscheinlichkeit (Filippi 1994, S.637 [77]).

Gd-aufnehmende Läsionen sind nach 14 Monaten Beobachtungsphase für die Entwicklung einer MS aus einem ersten demyelinisierenden Ereignis heraus spezifischer, aber weniger sensitiv als T₂-Läsionen (Miller 1996, S.7f. [104]).

4. Diskussion

4.1. Verbesserung der Validitätsbeurteilung durch Verwendung eines objektivierten Bewertungssystems

David Sackett et al. [135] haben 1996 mit den Fragen zur Beurteilung der Validität von Therapiestudien einen äußerst wichtigen Schritt auf dem Weg zu einem „bewussten, expliziten und vernünftigen Einsatz der gegenwärtig besten externen, wissenschaftlichen Evidenz für Entscheidungen in der medizinischen Versorgung individueller Patienten“ (Clade 2000, S.A3316 [66]) geschaffen. Seit DerSimonian und Kollegen 1982 vorschlugen, dass „... editors could greatly improve the reporting of clinical trials by providing authors with a list of items that they expected to be strictly reported“ (DerSimonian 1982, S.1336 [70]), hatte die Entwicklung ihren Lauf genommen. So entstand schließlich 1996 das CONSORT-Statement (Begg 1996 [63]). Etwa zwei bis drei Jahre vorher hatten Guyatt, Sackett und Kollegen die Fragen für die Beurteilung klinischer Studien entwickelt (Guyatt 1993 [86] und Guyatt 1994 [87]), die schließlich in einem Buch veröffentlicht wurden (Sackett, Harcourt Brace Publishers International 1996 [133] und Sackett 1999 [134]). Aufbauend auf diesen Fragen wenden vor allem die Cochrane Collaboration (www.cochrane.org [166]) sowie andere Gruppen in Deutschland und weltweit den Sackettschen Fragebogen (Abschnitt 2.4 dieser Arbeit; Guyatt 1993, S.2599 [86]) mit großem Erfolg bei der Beurteilung klinischer Studien an. Mit der Weiterentwicklung des Fragebogens zum objektivierten Bewertungsschema nach Hassler in der vorliegenden Arbeit (siehe Abschnitt 2.5) existiert nun ein System, in dem Therapiestudien noch objektiver als zuvor verglichen werden können. Diagramm 29 zeigt ein Flussdiagramm mit den wichtigsten Punkten zum Vorgehen bei einer Studienbewertung mit diesem Schema. Jedem der sechs Merkmale wird eine Punktzahl zugeordnet und die Verteilung der Punkte erfolgt entsprechend der Gewichtung der einzelnen Schritte. Die Möglichkeit einer Abwertung der jeweils untersuchten Publikation trägt dem Rechnung, dass Studien bei bestimmten Fehlern nicht als voll valide oder sogar als nicht valide eingestuft werden müssen. Da Patienteneigenschaften für jede Krankheit unterschiedlich wichtig sind, kann die Tabelle zur Bewertung der Übereinstimmung in Schritt 6 um die relevanten Punkte ergänzt und die vorgesehenen Eigenschaften gewichtet werden. Auch zusätzliche Aspekte einer Studie, die zu einer Einschränkung der Validität führen, werden berücksichtigt: Der Bewerter kann nach Abschluss der herkömmlichen Beurteilung in den Schritten 1 bis 6 und der daraus berechneten Gesamtpunktzahl weitere zuvor noch nicht involvierte negative Bereiche auflisten und daraus ableiten, welchen Validitätsgrad die jeweilige Studie aufgrund dieser Einschränkungen noch maximal unabhängig von der Gesamtpunktzahl erhalten kann. Um

das objektivierte Bewertungssystem nach Hassler zum Beispiel in Studien zur EBM-Qualität mit mehreren Bewertern einzusetzen, empfiehlt es sich, wie bei anderen diagnostischen Methoden, etwa der MRT, ein vorheriges Training durchzuführen, um die Inter- und Intrarater-Variabilität möglichst gering zu halten.

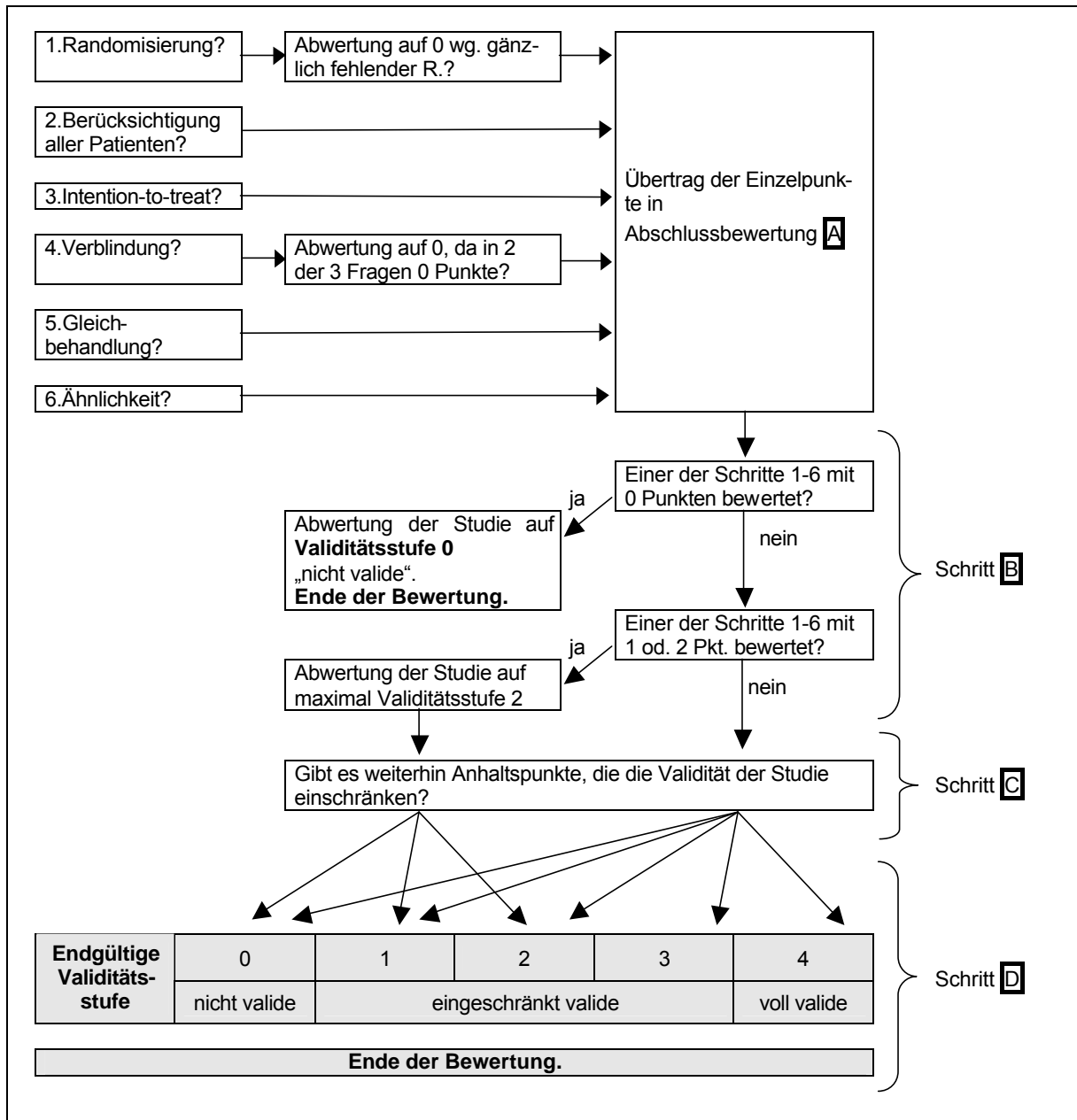


Diagramm 29: Flussdiagramm für das objektivierte EBM-Bewertungssystem nach Hassler.

Selbstverständlich ist jedoch auch dieses System nicht vollkommen sicher vor Fehlbeurteilungen. So existieren mit den Schritten 6 und **C** Teile, die nicht komplett objektiviert werden können und deren Bewertung auf der Erfahrung des Anwenders und der jeweiligen Krankheitsentität beruhen. Außerdem wurden zwar die Ergebnisse des Systems in der vorliegenden Arbeit mehrfach überprüft, dennoch gibt es derzeit kein anderes objektives System, das als Goldstandard zum Vergleich hätte dienen können. Somit ist trotz aller Sorgfalt bei der Erstellung der Schritte und der Einteilung der Validitätsgruppen nicht aus-

zuschließen, dass nach Überprüfung mit einem noch größeren Studienpool weitere Verbesserungen oder auch Änderungen der Validitätsgrenzen nötig sind. Schwierigkeiten können auch bei einem Vergleich mit den Validitätsgruppen anderer Schemata auftreten, wie zum Beispiel dem des Oxford Centre for Evidence-based Medicine (<http://minerva.minervation.com> [164]), da dort die Validität eher nach dem Aufbau als nach der exakten Durchführung einer Studie bestimmt wird (siehe Abbildungen 17 und 18 im Anhang). RCTs erhalten automatisch Validität 1 oder 2, die besten beiden Validitätsgruppen, während zum Beispiel Expertenmeinungen Validität 5 erhalten. In dem objektivierte Bewertungssystem nach Hassler gibt es jedoch die Möglichkeit, auch randomisierte kontrollierte Studien mit Validität 0, der schlechtesten Stufe, zu bewerten, wenn wichtige Punkte nicht beachtet wurden.

Trotz der beschriebenen Limitationen bietet das in dieser Arbeit neu erstellte objektivierte Bewertungssystem eine sehr gute Möglichkeit, Studien besser beurteilen zu können, ein Bild über die Qualität klinischer Studien zu erhalten und Handlungsvorschläge abzuleiten.

4.2. Erkenntnisse durch die Validitätsbewertung der eingeschlossenen Therapiestudien und Nutzen der Evidence-Based Medicine für die Durchführung einer wissenschaftlich gestützten objektiven Medizin

Um den Rahmen dieser Diskussion nicht zu sprengen, wird hier auf eine ausführliche Diskussion aller CONSORT-Kriterien ausdrücklich verzichtet. Lediglich auf die unter 3.1.2. genannten Bereiche, die besondere Wichtigkeit für die EBM-Bewertung haben, soll bei den jeweiligen EBM-Schritten kurz eingegangen werden.

In der vorliegenden Arbeit wird für eine Stichprobe von 38 randomisierten Therapiestudien zur Multiplen Sklerose gezeigt, wie häufig die Kriterien von CONSORT und EBM bereits erfüllt werden und wo in der Berichterstattung sowie in Validität, Wichtigkeit und Anwendbarkeit von Studienresultaten noch Probleme bestehen. Da nicht-randomisierte Studien als nicht valide zu gelten haben und somit für eine Bewertung nach EBM uninteressant sind (Sackett 1999, S.74 [134]), wurden in die vorliegende Arbeit lediglich randomisierte Studien eingeschlossen.

Für einige Aspekte der EBM, wie zum Beispiel die Auflistung der Patientendaten in einer Tabelle (EBM-Schritt 6) in 74% (28 von 38) aller bewerteten Studien oder auch für andere EBM-Forderungen ist ein positiver Trend zu verzeichnen. Das Gleiche gilt für die Ausführung verschiedener CONSORT-Teilbereiche. Es existieren mittlerweile einige Studien, in denen sowohl die CONSORT-Forderungen als auch die Schritte der EBM in befriedigender Weise umgesetzt werden. Die Richtung, in die sich die klinische Forschung bewegt,

stimmt in ihren Grundsätzen, auch wenn es noch viele Punkte gibt, an denen Verbesserungen dringend nötig sind. Denn die Erfüllung von CONSORT als Grundlage für die EBM ist besonders zu fordern, weil „die Qualität der Gesundheits-Versorgung ... ebenso wie die Qualität der Gesundheits-Forschung an der Transparenz der Darstellung gemessen werden (kann).“ (Porzsolt 2001, S.10 [125]). Je transparenter eine Studie ist, desto besser kann derjenige, der ihre Ergebnisse in der täglichen Praxis oder als Grundlage für weitere Forschungen heranzieht, beurteilen, ob die Resultate auch wirklich auf den speziellen Fall anwendbar sind. „Objektivität (ist) eine Voraussetzung für Reliabilität und diese wiederum eine Voraussetzung für Validität.“ (www.e-consulters.de [169]).

Schulz und Kollegen fanden zum Beispiel entsprechend dem Schritt 1 des Sackettschen EBM-Schemas, dass Studien, in denen die Patientenzuteilung inadäquat erfolgt oder unklar dokumentiert und deren Validität damit eingeschränkt war, den Behandlungserfolg im Durchschnitt überschätzten (Schulz 1995, S.410 [140]). Das beweist die Relevanz einer Studienanfertigung nach den Kriterien der EBM, insbesondere der Patientenzuteilung mittels Randomisierung.

Voraussetzung für die Bewertung der Zuweisung zu den Vergleichsgruppen in einer Studie ist nicht nur, dass diese überhaupt randomisiert durchgeführt wurde, sondern auch die Verwendung einer adäquaten Methode der Zufallszuteilung mit ausführlichen Angaben dazu einschließlich deren Besonderheiten, zum Beispiel Blockbildung, Stratifizierung oder Minimierung (CONSORT-Forderung Nummer 8), sowie zur Umsetzung der Zuteilung inklusive Geheimhaltung (CONSORT-Forderung Nummer 9). So wurde in nur 15 Studien (39%) die Randomisationsmethode (Forderung Nr.8) beschrieben (siehe Abschnitt 3.1.1.2.2.). In den übrigen Veröffentlichungen wurde lediglich erwähnt, dass überhaupt randomisiert worden war. Die Durchführung der Patientenzuteilung (Forderung Nr.9) wurde zu 61%, das heißt in 23 Studien, beschrieben, und 20 Publikationen (53%) enthielten Angaben, ob die Randomisierung geheimgehalten worden war (siehe Abschnitt 3.1.1.2.2.).

Es ist zu beachten, dass sich die Werte für CONSORT-Forderungen von denen der EBM-Bewertung unterscheiden, da für CONSORT alle ausführlichen Angaben, die sich nur in Schwesterstudien fanden, mit lediglich einem Punkt bewertet werden. Im Gegenzug wird für die CONSORT-Beurteilung wiederum als positiv gewertet, falls ein Fehlen des entsprechenden Merkmals ausführlich beschrieben wird. CONSORT fordert ja, wie schon im Abschnitt 2.3. dieser Arbeit erwähnt, primär eine genaue *Beschreibung* der Durchführung und erst im zweiten Schritt eine möglichst gute Ausführung der Studie. In der EBM-Bewertung kommt es jedoch nur darauf an herauszufinden, ob die geforderte Form der Durchführung, in diesem Fall also eine Randomisation, durchgeführt wurde oder nicht. Die

EBM-Ergebnisse (siehe Abschnitt 3.1.2.1.1.): 97% Randomisation durchgeführt, 90% Zufallszuteilung zum richtigen Zeitpunkt, aber nur in 61% Geheimhaltung inklusive der Studien, die lediglich „centrally randomised“ angaben.

Während 1979 und 1980 84% (56 Veröffentlichungen) aus 67 Publikationen von NEJM, JAMA, BMJ und Lancet eine Randomisierung angaben, berichteten lediglich 19% (13 RCTs) auch über die Methode der Zuteilung (DerSimonian 1982, S.1334 [69]). Um eine Vergleichbarkeit der Ergebnisse von DerSimonian und Kollegen mit den Ergebnissen der vorliegenden Arbeit, die ausschließlich randomisierte Studien untersuchte, zu ermöglichen, wurde der relative Anteil der Zuteilungsbeschreibung aus DerSimonian et al. 1982 [69] errechnet: 19% von 84% randomisierten Studien entspricht einem Prozentsatz von 22,6%, was bedeutet, dass nur in 22,6% aller untersuchten *randomisierten* Studien aus den Jahren 1979 und 1980 die Methode der Zufallszuteilung beschrieben war. Etwas besser stellten sich die Werte in gynäkologischen Studien von 1990 und 1991 dar, unter denen 32% (n=66) von 206 Veröffentlichungen die Randomisationsmethode angaben (Schulz 1994, S.126 [139]) Auch Altman und Doré konnten in nur 49% (n=39) von 80 randomisierten Studien aus dem Jahr 1987 Informationen über die Randomisationsmethode finden (Altman 1990, S.150 [50]). Williams und Davis schließlich, die eine Unterteilung der Beschreibung in vier Gruppen (good, fair, poor, none) vornahmen, fanden bei Studien der 1980er Jahre, publiziert in der Zeitschrift „Controlled Clinical Trials“, in 74% eine mangelhafte oder fehlende, hingegen in nur 26% (n=52) von 200 Veröffentlichungen eine gute oder ordentliche Beschreibung (Williams 1994, S.294ff. [157]). Dieser Prozentsatz hat sich auch im Jahr 2000 für MS-Studien nicht weiter verbessert und liegt bei 39% der in die Untersuchung eingeschlossenen Publikationen (siehe oben und Tabelle 27).

Studie	Publikationsjahr der untersuchten Veröffentlichungen	Gefundene Beschreibung der Randomisationsmethode in
DerSimonian 1982 [69]	1979 – 1982	22,6%
Altman 1990 [50]	1987	49%
Schulz 1994 [139]	1990 – 1991	32%
Williams 1994 [157]	1980er Jahre	26%
Aktuelle Untersuchungen für die vorliegende Arbeit	1997 – 2000	39%

Tabelle 27: Gefundene Häufigkeit der Beschreibung der Randomisationsmethode in früheren Studien im Vergleich mit den Ergebnissen der vorliegenden Arbeit.

Insgesamt zeigt sich, dass der Leser einer Veröffentlichung mit Ergebnissen zwischen 22,6% bis 49% noch viel zu selten über die Zuteilungsmethode informiert ist, um daraus ein Urteil über die Wirksamkeit der Randomisation fällen zu können. Weitere Verbesserungen auf diesem Gebiet sind dringend nötig.

Neben der Frage, ob überhaupt randomisiert wurde, ist für die Erfüllung von Schritt 1 des EBM-Schemas die Geheimhaltung der Zuteilung relevant. In den EBM-Auswertungen für die vorliegende Arbeit konnte diese in lediglich zwölf Studien (32%) und deren Nebenbe-

richterstattungen sicher gefunden werden. Weitere elf Publikationen (29%) erwähnten eine zentrale Randomisation, was zumindest auf eine mögliche Geheimhaltung schließen lässt. Interessant ist, dass in einer anonymen Befragung mehr als die Hälfte aller an Studien beteiligten Ärzte zugaben, dass die Geheimhaltung in mindestens einer Studie, an der die Befragten mitgewirkt hatten, nicht gewährleistet war (Schulz 1996, S.597 [141]). Dies bedeutet jedoch nicht, dass das auch für über die Hälfte aller Studien zutrifft, da bei dieser Umfrage viele der Befragten an mehreren RCTs beteiligt waren. Dennoch zeigt sich darin einer der Mängel, die in klinischen Studien auch heutzutage noch viel zu häufig vorhanden sind. Die Umfrageresultate lassen sich vereinbaren mit den Untersuchungsergebnissen in der vorliegenden Arbeit, wobei eine fehlende Geheimhaltung oder zumindest mangelnde Beschreibung in mindestens 39% aller Studien festgestellt werden musste (siehe Abschnitt 3.1.2.1.1.). Im Prinzip existieren drei Möglichkeiten, einem Bekanntwerden der Patientenzuteilung entgegenzutreten:

1. Zuteilung mittels Telefon durch ein zentrales Randomisationszentrum;
2. Zuteilung durch die Krankenhausapotheke mittels kodierter Medikamentenverpackungen, deren Inhalt nur dem Apothekenpersonal bekannt ist;
3. Zuteilung durch ein sicheres System nummerierter versiegelter Umschläge; (Altman 1991, S.1481 [51]).

Zusammenfassend lässt sich also für den EBM-Schritt 1 – Randomisation inklusive einer geeigneten Randomisationsmethode und Geheimhaltung – festhalten, dass zwar viele klinische Studien der Forderung nach einer Randomisation gerecht werden, aber die Methoden inklusive deren Beschreibung (siehe Tabelle 27) und die Geheimhaltung der Zufallszuteilung häufig noch als verbesserungsbedürftig beurteilt werden müssen. Dies wird bereits postuliert durch Studien von DerSimonian 1982 [69], Altman 1990 [50], Schulz 1994 [139] und Williams 1994 [157] und konnte in den Untersuchungen für die vorliegende Arbeit deutlich bestätigt werden.

Für EBM-Schritt 2 – Einschluss aller Patienten – konnten Schulz und Kollegen in Studien, die Patienten nach der Randomisation ausgeschlossen hatten, keine Überbewertung des Behandlungseffekts finden (Schulz 1995, S.410 [140]). Somit würde die Aussage von Sackett und Kollegen, die EBM-Schritte 1 (Randomisation) und 2 (Einschluss aller Patienten) seien die wichtigsten bei der Beurteilung einer Studie (Sackett 1999, S.76 [134]), zumindest teilweise widerlegt. Da außerdem von Schulz et al. nachgewiesen werden konnte, dass die Verblindung eine wesentliche Rolle für Studienergebnisse spielt (Schulz 1995, S.410 [140]), muss vor allem dieser Schritt – im Gegensatz zu den Ausführungen von Sackett et al. – ebenfalls als besonders wichtig betrachtet werden. Den Einschluss aller

Patienten betreffend wird von Schulz und Kollegen jedoch auch angemerkt, dass dieses Ergebnis dadurch zustande gekommen sein könnte, dass nur ein Teil der in Schulz 1995 eingeschlossenen Studien auf diesen Zusammenhang hin untersucht worden war (Schulz 1995, S.410 [140]). Allgemein wäre eine weitere mögliche Erklärung, dass manche Autoren unter Umständen über Patienten, die nachträglich aus der Studie ausgeschlossen werden, von vornherein nicht berichten. Somit wird in deren Studien scheinbar die selbe Anzahl an Studienteilnehmern randomisiert, die später auch zur Auswertung gelangt, in Wahrheit sind aber einige Patienten verlorengegangen (Schulz 1996, S.597 [141]). Ein derartiges Vorgehen ließe sich natürlich selbst durch intensives Literaturstudium nicht nachweisen und kann deshalb in einer EBM-Bewertung leider nicht erfasst werden. Es muss, wie in vielen anderen Bereichen auch, an die Ehrlichkeit der Autoren appelliert werden, wenn es darum geht, valide und wichtige Ergebnisse zu veröffentlichen. Ein Verhalten wie von Schulz und Kollegen 1996 beschrieben wäre äußerst schädigend für die gesamte Wissenschaft und darf in der modernen Forschung nicht mehr vorkommen. Trotz dieser Bedenken ist als positiv zu werten, dass in immerhin 50% der 38 für die vorliegende Arbeit untersuchten Studien mehr als 90% aller anfangs randomisierten Patienten in die Endauswertung eingeschlossen worden waren. Leider wertete jedoch immer noch ein knappes Fünftel (18%) weniger als 70% aller aufgenommenen Studienteilnehmer aus (siehe Abschnitt 3.1.2.1.2.).

CONSORT geht hier noch einen ganzen Schritt weiter. Es wird in Abschnitt Nummer 13 der Checkliste nicht nur die Nennung der Anzahl der Patienten, die am Ende zur Auswertung gelangten, gefordert, sondern auch eine Beschreibung des Verlaufs der Patienten durch eine Studie. Dies ist am besten mittels eines Flussdiagramms zu bewerkstelligen, das in 76% aller eingeschlossenen Publikationen dieser Arbeit zu finden war (siehe Abschnitt 3.1.1.2.3.). Egger und Kollegen konnten ein Flussdiagramm in 139 von 270 RCTs (51,5%) aus dem Jahr 1998 finden, allerdings mit großen Unterschieden zwischen den einzelnen untersuchten Journals (Egger 2001, S.1996 [71]). Zusätzlich fordert die CONSORT-Gruppe in Nummer 16 eine exakte Angabe der Patientenzahl, die in die jeweilige Analyse eingeschlossen wurde. Diese Zahl konnte in 36 Publikationen (95%) mindestens in knapper Form nachgewiesen werden (siehe Abschnitt 3.1.1.2.3.). Damit waren die Voraussetzungen für eine Bewertung des EBM-Schritts 2 erfüllt. Die oben genannten Werte für den Prozentsatz der ausgewerteten Patienten in den einzelnen Studien sind somit tatsächlich gefundene Häufigkeiten und nicht, wie teilweise in anderen Schritten der EBM-Bewertung, bedingt durch Abwertungen aufgrund mangelnder Transparenz.

Zwar konnte für einen zu geringen Patienteneinschluss keine signifikante Veränderung der Studienergebnisse nachgewiesen werden (Schulz 1995, S.410 [140]), solange aber auch kein gesichertes gegenteiliges Ergebnis vorliegt, sollte im Rahmen jeder Studie der

Versuch unternommen werden, möglichst viele der anfangs eingeschlossenen Patienten am Ende in die Auswertung zu bringen, wobei diese nach Intention-to-treat erfolgen sollte. Nicht nur, um den CONSORT-Forderungen zu genügen, sondern da dies auch Einfluss auf die EBM-Bewertung nimmt, sollte für eine bessere Transparenz außerdem jede klinische Studie ein ausführliches Flussdiagramm enthalten, das die Bewegung der Patienten in der Studie exakt darstellt. Dieses Diagramm sollte im Einzelnen folgende Punkte enthalten:

1. Die Anzahl der randomisierten Studienteilnehmer
2. Die Anzahl der Patienten, die die vorgesehene Studienmedikation erhielten
3. Die Anzahl der Teilnehmer mit vollständig durchgeführtem Studienprotokoll
4. Die Anzahl derjenigen, die schließlich auf das primäre Zielkriterium untersucht wurden. (Altman 2001, S.677-679 [48])

Ein derartiges Diagramm ist nämlich assoziiert mit „... improved quality of reporting of randomized controlled trials“ (Egger 2001, S.1999 [71]) und könnte, nach Altman 2001 [48] (S.666) wie in Diagramm 30 dargestellt werden. Gute Beispiele für Flussdiagramme finden sich auch in den Abbildungen 11 und 12 im Anhang.

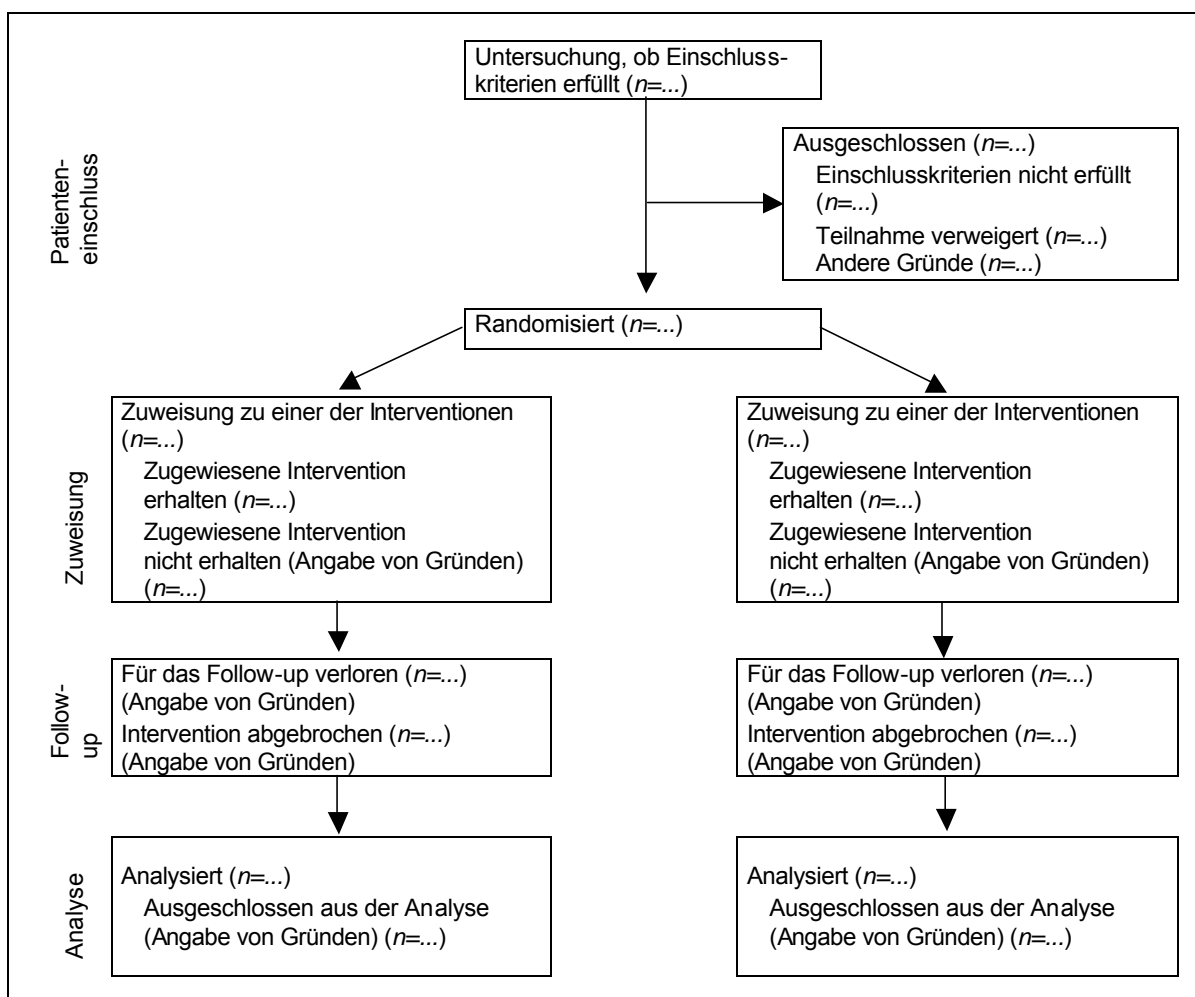


Diagramm 30: Flussdiagramm für den Verlauf von Studienteilnehmern während einer Studie. (vgl. Altman 2001, S.666 [48])

Es folgt der nächste wichtige Aspekt des EBM-Bewertungsschemas, Schritt 3 in der Bewertung nach Sackett – die Frage nach einer durchgeführten Intention-to-treat-Analyse. Hollis und Campbell fanden in ihrer Untersuchung dazu, dass nur 48% (119 Publikationen) in einer Überprüfung von 249 Veröffentlichungen eine Intention-to-treat-Analyse erwähnten (Hollis 1999, S.670f. [90]). Ruiz-Canela und Kollegen konnten die Erwähnung einer ITT in 47,7% (290 von 608 Veröffentlichungen) sichern (Ruiz-Canela 2000, S.1007 [132]), und in den Untersuchungen für die vorliegende Arbeit konnte die Erwähnung einer Auswertung nach ITT in 58% (22 von 38 Publikationen) nachgewiesen werden (siehe Abschnitt 3.1.2.1.3.).

Studie	Publikationsjahr der untersuchten Veröffentlichungen	Gefundene Beschreibung einer ITT-Analyse in
Hollis 1999 [90]	1997	48%
Ruiz-Canela 2000 [132]	1993 – 1995	47,7%
<i>Aktuelle Untersuchungen für die vorliegende Arbeit (vgl. Bewertung nach EBM-Kriterien*)</i>	1997 – 2000	58%
<i>Aktuelle Untersuchungen für die vorliegende Arbeit (nur Primärliteratur, entsprechend Bewertung nach CONSORT-Kriterien*)</i>	1997 – 2000	43%

Tabelle 28: Gefundene Häufigkeit der Beschreibung einer ITT-Analyse in früheren Studien im Vergleich mit den Ergebnissen der vorliegenden Arbeit. *)siehe Text.

Die leichte Steigerung der Prozentzahl lässt sich vereinbaren mit den Erkenntnissen von Ruiz-Canela und Kollegen, die herausfanden, dass die Anzahl an Studien mit Bericht einer ITT-Analyse zwischen 1993 und 1995 gering angestiegen war (Ruiz-Canela 2000, S.1007 [132]). Da die für die vorliegende Arbeit eingeschlossenen Studien zwischen 1997 und 2000 publiziert wurden, wäre ein weiterer Anstieg im Vergleich zu Ruiz-Canela 2000 für diesen Zeitraum denkbar. Die Änderung ist jedoch nicht signifikant (CHI-Quadrat=2,13), weswegen dieses Ergebnis nur als möglicher leichter Trend betrachtet werden darf. Außerdem ist es denkbar, dass Ruiz-Canela und Hollis jeweils nicht in Sekundärberichten nach der Angabe einer ITT-Analyse forschten. Um auch nur die Ergebnisse der Primärliteratur vergleichen zu können, sind in Tabelle 28 zusätzlich die Resultate aus der Auswertung des zweiten Teils von CONSORT-Forderung Nummer 16 aufgelistet (CONSORT-Werte wurden ja nur dann als voll erfüllt betrachtet, wenn die Beschreibung in der Primärliteratur vorhanden war). Zur Bewertung in Schritt 3 nach Sackett bietet diese Forderung nämlich die beste Voraussetzung: Die Angabe, ob es sich bei der Auswertung um eine ITT-Analyse handelte. In diesem Punkt stimmen das CONSORT-Statement (Altman 2001 [48]) und Sackett et al. 1999 [134] im Prinzip direkt überein. Der gewissenhafte Autor kann dadurch dem kritischen Leser mit Erfüllung der CONSORT-Forderung die nötigen Informationen für eine positive Bewertung der Frage nach einer ITT-Analyse bei der EBM-Auswertung an die Hand geben. In der Studienevaluation der vorliegenden Arbeit wurden jedoch, wie bereits erwähnt, im Unterschied zur EBM-Bewertung, für CONSORT nur die Angaben als erfüllt angesehen, die in der untersuchten

Publikation vorhanden waren. Erläuterungen in Schwesterstudien oder Zweitberichten wurden als zum Teil erfüllt betrachtet. So war eine direkte Angabe einer ITT-Analyse lediglich in 16 Studien (43%) zu finden (siehe Abschnitt 3.1.1.2.3.), womit das Ergebnis sogar leicht unter dem von Hollis und Campbell sowie Ruiz-Canela et al. liegt.

Zusammenfassend kann man also sagen, dass bis heute nur etwa die Hälfte aller Publikationen eine Auswertung nach ITT zumindest erwähnt, wobei sich der Leser dabei noch nicht auf deren tatsächliche Durchführung verlassen darf. Um aber die Randomisierung aufrecht zu erhalten sowie Non-Compliance und Abweichungen vom Studienprotokoll zu berücksichtigen (Hollis 1999, S.670f. [90]), ist eine ITT-Analyse zwingend erforderlich. Gerade Patienten, die eine Behandlung abbrechen, sind „... oft schwerer betroffen ... oder (haben) vermeintlich weniger profitiert (Damit) könnte ihr Ausscheiden zu Verzerrungen der Ergebnisse führen“ (Gold 2000, S.61 [84]). Eine Auswertung die nicht nach ITT erfolgt, zerstört die vorher durchgeführte Randomisation (Altman 1996, S.596 [49]).

Für Schritt 4 – Verblindung von Patienten und Ärzten – konnten Schulz und Kollegen 1995 aufzeigen, dass Studien, in denen keine Doppelblindheit vorliegt, den Erfolg der experimentellen Therapie tendenziell überbewerten (Schulz 1995, S.410 [140]). Darin zeigt sich die Wichtigkeit der Verblindung für die Gültigkeit der Ergebnisse.

In der vorliegenden Arbeit konnte bei der Auswertung der eingeschlossenen 38 Publikationen am häufigsten eine Verblindung der Diagnostiker sicher festgestellt werden (65%), am zweithäufigsten eine Verblindung der Patienten (63%) und nur in 21% der bewerteten Veröffentlichungen eine Verblindung der Therapeuten. Zusätzlich war in 24% aller Studien die Verblindung der Patienten nicht eindeutig, außerdem in 32% die Verblindung der Diagnostiker möglich, aber nicht sicher und sogar in 42% der Therapeuten das Kriterium nur unter Umständen erfüllt (siehe Abschnitt 2.4.1.4.). Nach CONSORT hingegen war bei der Durchsicht der Studien in insgesamt 89% die Verblindung der Patienten mindestens knapp beschrieben, sogar in 97% die Verblindung der Diagnostiker, jedoch in nur 50% die Verblindung der Therapeuten (siehe Abschnitt 3.1.1.2.2.). Zu den Unterschieden in der Auswertung von CONSORT Nummer 11 und EBM-Schritt 4 kommt es aus verschiedenen Gründen. Einerseits legt CONSORT besonders Wert auf eine gute Berichterstattung, und weniger darauf, ob eine Verblindung tatsächlich stattfand oder nicht. Andererseits achtet EBM wiederum mehr auf eine in Wahrheit stattgefundene Verblindung. So wurden Studien, in denen außer der doppelblinden Durchführung weder positive noch negative weitere Angaben vorhanden waren, trotzdem mit zwei Punkten in der EBM-Auswertung versehen, während es für derartig knappe Angaben in der CONSORT-Bewertung nur maximal einen Punkt gab. Im Gegensatz konnte eine Studie, in der festgehalten ist, dass für eine Gruppe keine Verblindung vorlag, von CONSORT zwei Punkte erhalten, während EBM

natürlich null oder maximal einen Punkt vergeben muss. Als Beispiel lässt sich hier Cook 1997 nennen, wo „... a significant difference ... in patient perception as to treatment received in the two groups“ (Cook 1997, S.178 [3]) gefunden wurde. In der EBM-Auswertung wurde hierfür immerhin ein Punkt vergeben, da von vornherein eine Verblindung vorhanden war, die nur im Verlauf der Studie zum Teil vermindert wurde. Nach CONSORT wurden, wegen der sehr guten Beschreibung, zwei Punkte zugewiesen.

Eine Gegenüberstellung der Auswertungsdaten mit der Verblindung von Patienten im Jahr 1980 bietet DerSimonian et al. 1982 [69]. Dort wird beschrieben, dass bei 55% aller Studien (37 von 67 Publikationen) die Patienten zuverlässig verblindet waren und bei weiteren 12% (8 von 67 Veröffentlichungen) dieses Kriterium nicht ganz sicher erfüllt war. Zum Vergleich: Bei der Bewertung für die vorliegende Arbeit konnten in $63\% + 24\% = 87\%$ aller Studien die Patienten als eindeutig oder nicht ganz sicher verblindet erkannt werden (siehe Abschnitt 2.4.1.4.) – 1980 waren es $55\% + 12\% = 67\%$ (DerSimonian 1982, S.1334 [69]). Auch eine Gegenüberstellung der Daten für die diagnostisch tätigen Ärzte ist möglich: Bei DerSimonian et al. finden sich in 30% (20 von 67 Studien) eine sichere und in weiteren 27% (18 von 67 RCTs) eine mögliche Verblindung, zusammengenommen also in 57%. Bei der Bewertung für die vorliegende Arbeit waren es $65\% + 32\% = 97\%$, die weit überwiegende Mehrheit. Allerdings muss ein Vergleich der Zahlen beider Studien mit Vorsicht erfolgen, da es sich bei den jeweils eingeschlossenen Publikationen um Untersuchungen mit verschiedenen Themen handelt und die Ergebnisse hier nur für MS-Therapiestudien gelten. Dennoch ist über die Jahre vor allem für Diagnostiker, aber auch Patienten, ein Trend ($\text{CHI-Quadrat}(\text{Diagnostiker})=45,17$ ($p<0.01$); $\text{CHI-Quadrat}(\text{Patienten})=11,29$ ($p<0,01$)) in Richtung bessere Verblindung und damit erhöhte Qualität zu verzeichnen, auch wenn die Berichterstattung und die Sicherung der Verblindung insgesamt bei weitem noch nicht ausreicht. Wie auch für die Randomisation ist es wichtig, die Methoden für eine Aufrechterhaltung der Verblindung zu publizieren, damit sich der Leser ein Bild davon machen kann, wie resistent eine Studie gegenüber der Gefahr einer Aufhebung der Verblindung ist (DerSimonian 1982, S.1334 [69]). Schließlich beanspruchen alle für die vorliegende Arbeit eingeschlossenen Publikationen das Kriterium der Doppelverblindung für sich, was nicht in jeder Veröffentlichung nachzuvollziehen ist. Eigentlich kann man sich einer Verblindung nur dann wirklich sicher sein, wenn diese auch mittels eines Fragebogens oder ähnlich evaluiert wurde. Ansonsten ist ein Nachlassen der Verblindung durch Nebenwirkungen, Öffentlichwerden des Randomisationscodes oder ähnliches nicht sicher auszuschließen. Daher ist ein Bericht über die Ergebnisse einer derartigen Untersuchung zur Qualitätssteigerung dringend anzuraten.

Eine Überprüfung der Abhängigkeit der Behandlungsergebnisse von einer fehlenden Verblindung wie bei Schulz et al. 1995 [140] war aufgrund der relativ geringen Anzahl an

Publikationen und deren Divergenz der darin erforschten Therapieoptionen in der vorliegenden Arbeit leider nicht möglich.

Schritt 5 in der EBM-Bewertung von Therapiestudien stellt die Frage nach der Gleichbehandlung der Studienpatienten. Das bedeutet, dass außer der Studienmedikation alle Gruppen exakt identisch betreut werden müssen. Waren das Studienprotokoll und die Studienorganisation (entsprechend CONSORT-Forderung 3b) auf eine Gleichbehandlung ausgelegt und die Studie randomisiert, kann, falls keine Fehler bei der Durchführung auftraten, von einer Gleichbehandlung ausgegangen werden. Wichtig für die Gleichbehandlung sind auch die Orte der Studiendurchführung (ebenfalls in CONSORT Nummer 3b), da unterschiedliche Krankenhäuser verschiedene Organisationsstrukturen, Erfahrungen und Ressourcen haben. Außerdem spielen unter anderem Klima, geografische Faktoren und das soziokulturelle Milieu eine Rolle. All das kann zusätzlich Einfluss auf die Wirksamkeit einer Studienmedikation haben (Altman 2001, S.668 [48]). Keine der eingeschlossenen Studien berichtete jedoch ausführlich genug über die Lokalitäten der Durchführung (siehe Abschnitt 3.1.1.2.2.). Auch CONSORT Nummer 4 schafft mit der Forderung nach Beschreibung der geplanten Interventionen und deren Durchführung Voraussetzungen für eine Beurteilung der Gleichbehandlung. In den bewerteten Publikationen konnte dieses Kriterium in 55% als erfüllt angesehen werden. Somit bieten die CONSORT-Forderungen 3b und 4 mit weiteren Faktoren zusammen die Voraussetzung, um in diesem Punkt eine Bewertung nach EBM vornehmen zu können. Schon deshalb, weil bei den in die Evaluationen der vorliegenden Arbeit eingeschlossenen Publikationen häufig zur Behandlung von Exazerbationen zusätzliche Medikamente erlaubt waren, kann die Gleichbehandlung nach EBM meistens nicht als vollständig, sondern nur als zum Teil vorhanden angesehen werden. Dies war in 36 Publikationen (95%) der Fall. Wie unter 3.1.2.1.5. bereits ausführlich berichtet, behandelten außerdem zwei Studien (5%) die Vergleichsgruppen eher nicht bzw. zu einem Teil nicht gleich (Optic Neuritis Study Group 1997, S.1405 [29], Koch-Henriksen 2000, S.173 [16]).

Sackett et al. 1999 erwähnen zwar diesen Schritt der EBM-Bewertung, betrachten ihn jedoch als zweitrangig, indem dazu keinerlei zusätzliche Erklärungen gegeben werden (Sackett 1999, S.76 [134]). Auch in der Literatur wird die Gleichbehandlung kaum erwähnt. Dafür könnten möglicherweise folgende zwei Gründe verantwortlich sein:

1. Die Gleichbehandlung wird von den meisten Autoren als selbstverständlich angesehen und daher nicht weiter beachtet.
2. Mit einem entsprechenden Studienprotokoll, das durch Gegenüberstellung von Placebo- und Verumgruppe, deren einziger Unterschied die experimentelle Therapie ist, scheinbar Gleichbehandlung gewährleistet, wird normalerweise zusammen

mit einer randomisierten Zuordnung grundsätzlich für eine Erfüllung dieses Schritts gesorgt. Auf eine Ähnlichkeit müsste dann theoretisch nicht noch einmal verstärkt eingegangen werden.

Dennoch werden Fehler wie bei Koch-Henriksen 2000 [16] oder bei der Optic Neuritis Study Group 1997 [29] nur durch zusätzliche Überprüfung der Frage der Gleichbehandlung aufgedeckt. In Zukunft sollte also mehr Augenmerk auch auf den fünften Schritt der EBM-Bewertung nach Sackett gelegt werden, da dieser ebenfalls, wenn auch in der Rangfolge hinter Randomisierung und Verblindung, eine wichtige Rolle spielt. Allerdings existieren soweit bekannt in der Literatur weder zu der Gleichbehandlung noch zu der Ähnlichkeit der Therapiegruppen genaue Untersuchungen, in welchem Ausmaß eine Nichterfüllung der Kriterien zu einer Beeinflussung der Studienergebnisse führt. Sicher ist jedoch, dass man Wirkungen der experimentellen Therapie dieser nicht mehr zuordnen kann, wenn alleine eine der beiden Gruppen zusätzlich weitere Therapien erhält, die der anderen vorenthalten bleiben.

Auch bei der Ähnlichkeit der Vergleichsgruppen gibt es keine auf Zahlen gestützten allgemeinen Erkenntnisse, ab welchen Unterschieden in den Gruppen mit einem Bias zu rechnen ist. Selbstverständlich ist der Einfluss der Faktoren stark von der jeweils untersuchten Krankheit und deren Risikofaktoren abhängig. Jedoch ist offensichtlich, dass eine Studie spätestens dann nicht mehr als valide angesehen werden kann, wenn die für eine bestimmte Krankheit wichtigen Risikofaktoren signifikant unterschiedlich zwischen den Therapiegruppen verteilt sind. In der Regel führt allerdings eine gut durchgeführte Randomisierung zu einer Gleichverteilung der bekannten und der unbekannt Risiken. Eine Zufallszuteilung bietet aber keine Garantie für eine Gleichheit der Merkmale (Altman 2001, S.680 [48]). Das bestätigt sich auch in der Auswertung der bei der Überprüfung gefundenen Daten. In über 70% (n=20) sind die Vergleichsgruppen der bewerteten Studien ähnlich, bezogen auf die 74% (n=28) aller Publikationen, in denen eine Tabelle mit Auflistung der relevanten Patientendaten, wie von CONSORT Nummer 15 gefordert, vorhanden war. Interessanterweise konnte keine der Veröffentlichungen ohne eine derartige Tabelle eine mindestens weitgehende Übereinstimmung beweisen, was nicht bedeutet, dass sie nicht vorhanden war. Es fehlten lediglich die Angaben, um diesen Studien ein besseres Ergebnis in EBM-Schritt Nummer 6 zuzusprechen. Daher muss den Autoren jeder randomisierten Studie empfohlen werden, wenigstens die wichtigsten Daten in Form einer Tabelle oder, als Minimalforderung, übersichtlich in den Text zu integrieren (Altman 2001, S.680 [48]). Als gutes Beispiel kann dazu Table 1 aus The Lenercept Multiple Sclerosis Study Group 1999 [41] (S. 460) gelten (siehe Abbildung 13), wohingegen Table 3 aus Jacobs 1995 [93] (S.126) wegen fehlender Gegenüberstellung der Vergleichsgruppen als nicht suffizient anzusehen ist (siehe Abbildung 14 im Anhang).

Die aus den Bewertungsergebnissen (siehe Abschnitt 3.1.2.1.6.) zu ziehende Schlussfolgerung ist, dass in der überwiegenden Mehrzahl randomisierter kontrollierter Studien die Patientengruppen aufgrund der Zufallszuteilung ähnlich sind oder sich nur relativ gering unterscheiden. Meistens fehlt der Bericht. Es ist von großem Nachteil für die Validität einer Veröffentlichung, wenn eine Ähnlichkeit der Patientengruppen aufgrund mangelnder Transparenz nicht nachvollziehbar ist. Im Text werden so gut wie nie alle notwendigen Daten aufgeführt, da die Autoren von Veröffentlichungen verständlicherweise Schwierigkeiten haben, dort so systematisch wie in einer Tabelle vorzugehen, und selbst bei einem Vorliegen der Informationen sind die Gruppen für den Leser im Text wesentlich schwieriger zu vergleichen. Dies ist einmal mehr ein interessantes Beispiel für die Notwendigkeit der Erfüllung der CONSORT-Forderungen (hier speziell der Forderung nach einer Patiententabelle, Nummer 15 auf der CONSORT-Checkliste) als Basis für eine gute Bewertung nach EBM-Kriterien.

Generell gilt: Die EBM soll den anwendenden Arzt befähigen, „... eine möglichst hochstehende Medizin nach dem aktuellen Wissensstand zu praktizieren, ohne selbst jeden Tag viele Stunden neueste Untersuchungen zu studieren, zu prüfen und daraus praktische Handlungsanleitungen zu gewinnen“ (Clade 2000, S.A3316 [66]). Bei allen Ergebnissen der EBM-Bewertungen sollte jedoch nicht vergessen werden, dass eine adäquate Berichterstattung, den Forderungen des CONSORT-Statements folgend, erst die Voraussetzungen für eine Bewertung nach EBM liefert. Die dabei evaluierte interne Validität, die die in einer Studie untersuchte Population betrifft, ist schließlich die Basis für die externe Validität, also die Umsetzung von Studienergebnissen in der Gesamtpopulation. Bei nicht vorhandener oder eingeschränkter interner Validität ist auch die externe Validität von vornherein vermindert.

4.3. Vorschläge für die Verbesserung der Wichtigkeit und Anwendbarkeit von Therapiestudien

Für die Auswertung der Wichtigkeit der eingeschlossenen MS-Therapiestudien wurden die drei in der EBM etablierten Kennzahlen „Relative Risikoreduktion“ (RRR), „Absolute Risikoreduktion“ (ARR) und „Number-needed-to-treat“ (NNT) verwendet (siehe Abschnitt 2.4.2.). Die EBM verlangt ein Vorgehen in folgender Reihenfolge:

1. Finden der klinisch nützlichsten Aussage
2. Errechnung der Kennzahlen RRR, ARR und NNT

3. Gegenüberstellung der Ergebnisse mit den Ergebnissen anderer Therapien bei anderen Erkrankungen, um Vergleichswerte über den Nutzen einer Behandlung zu erhalten. (Sackett 1999, S.105-111 [134])

Häufig kann jedoch das Vorgehen nicht wie gefordert erfolgen, denn viele Studien geben nicht die für eine Berechnung notwendigen Werte an. Somit ist der Leser gezwungen, auf andere Zahlen auszuweichen, um eine Aussage zu der Wichtigkeit von Studienresultaten machen zu können. Allerdings entsprechen diese Werte oftmals nicht der klinisch nützlichsten Aussage. Dennoch ist es besser, überhaupt Kennzahlen zu erhalten als auf diese Auswertung ganz zu verzichten, da nur so Wichtigkeitsdaten mit Vergleichscharakter gewonnen werden können. Es ist sinnvoll, bei einer gleichen Aussage, zum Beispiel die Anzahl an Patienten ohne Exazerbationen nach einem bestimmten Zeitintervall, in zwei Studien mit unterschiedlicher experimenteller Intervention anfangs die jeweiligen NNTs dieser Studien miteinander zu vergleichen und erst dann eine Gegenüberstellung mit fachfremden Studien vorzunehmen. So kann man zunächst abschätzen, welche der beiden untersuchten Therapieoptionen die günstigere für das gewünschte Zielkriterium und anschließend, ob dieser Wert generell eher als gut oder weniger gut einzustufen ist.

Es existiert jedoch das Problem, dass die NNT-Werte einzelner Studien in der Regel nicht miteinander vergleichbar sind, weil die Zielkriterien nur selten übereinstimmen. Eine wenn auch sicherlich nur schwer zu realisierende Möglichkeit zur Behebung dieses Problems wäre, dass die für die verschiedenen Krankheiten zuständigen einzelnen Fachgesellschaften zum Beispiel fünf Zielkriterien definierten, die als Empfehlung für zukünftige Studien dienen könnten. Jeder Studie auf dem entsprechenden Gebiet wäre dann empfohlen, dieses Kriterium, wo immer möglich, wenigstens als eines von mehreren Outcomes anzuwenden. Damit wäre das Problem zumindest insofern behoben, dass in Reviews nicht mehr „Äpfel mit Birnen“ verglichen werden müssten.

Außerdem ist es für die NNT-Berechnung notwendig, dass ein Zielkriterium zweiseitig angelegt und auf die Patienten bezogen ist. Um eine Berechnung vornehmen zu können, muss eine Ja-Nein-Entscheidung möglich sein, wie beispielsweise die Anzahl der Patienten mit Exazerbation versus diejenigen ohne Exazerbation, aufgeschlüsselt nach Behandlungsgruppe, wie zum Beispiel in Tabelle 4 von Tubridy 1999 [43] (S.470). Damit lässt sich der CHI-Quadrat-Test und die NNT-Berechnung mittels Vierfeldertafel durchführen. Dies funktioniert nicht, wenn ein Kriterium nur als Zahl angegeben ist, beispielsweise die Rate neuer Läsionen pro Patient nach einem bestimmten Zeitintervall (vgl. Kastrukoff 1999, S.355, Tabelle2 [15]). Tabelle 12 der vorliegenden Arbeit enthält eine Auswahl der errechneten Kennzahlen aus den mindestens mit Validitätsstufe 2 bewerteten Studien. Diese wurden für die Evaluation des Therapieerfolgs im Review des Abschnitts 3.2. eingesetzt.

4.4. Vergleich der Therapiemöglichkeiten bei Multipler Sklerose

Aus den vielzähligen untersuchten Therapiemöglichkeiten der Multiplen Sklerose kristallisieren sich nach den in die vorliegende Arbeit eingeschlossenen Studien und dem Review von Clegg 2000 [67] drei Gruppen mit fließenden Übergängen heraus. Die erste Gruppe umfasst die Stoffe, die eine Wirksamkeit entweder bereits belegt haben oder deren Behandlungswirkung äußerst vielversprechend scheint, aber noch in weiteren Studien verifiziert werden muss. Dazu gehören die Beta-Interferone, Azathioprin und die Lymphknotenbestrahlung in Kombination mit Kortikosteroiden. Als in der Übergangsphase zwischen den Gruppen 1 und 2 befindlich müssen Mitoxantron und Glatirameracetat, das in Deutschland bereits zur MS-Behandlung zugelassen ist, angesehen werden. Der zweite Typ Medikamente scheint in Teilbereichen wirksam, erweist sich aber entweder bei anderen Zielkriterien als nicht sinnvoll, oder die Ergebnisse können mangels Validität der zugrundeliegenden Studien nicht bestätigt werden. Dies trifft für Cladribin, Methotrexat, Immunglobulin G und Glukokortikoide zu. Schließlich existiert die dritte Art der Behandlungsoptionen, die ihre Wirksamkeit bei Multipler Sklerose nicht bestätigen konnten oder zu toxisch sind: Lenercept, Sulfasalazin, AG284, Antegren, Cyclophosphamid und Cyclosporin A.

Auch wenn für das Review der vorliegenden Arbeit nicht valide Studien ausgeschlossen wurden, ist zu bedenken, dass sowohl die für die eigenen Berechnungen als auch die in Clegg 2000 [67] verwendeten Studien immer noch von heterogener Qualität sind und oftmals gut gemachte Publikationen nur relativ kleine Patientenzahlen aufweisen. Dabei sind die wichtigsten Probleme in Veröffentlichungen mit eingeschränkter Validität laut Bryant und Kollegen eine unzureichende Verblindung, schlechte Beschreibung von Studienabbrüchen sowie unvollständige Berichterstattung über die Resultate und die damit zusammenhängenden Informationen (Bryant 2001, S.578 [64]). Außerdem unterscheiden sich die Applikationsarten und -häufigkeiten, Patientengruppen und Zielkriterien der einzelnen Studien, womit eine Vergleichbarkeit der Therapieformen nur noch bedingt gegeben ist. Als eine Besonderheit bei Multipler Sklerose kommt hinzu, dass sich über die Jahre ein Wechsel der Terminologie vollzogen hat und nun die ursprünglich nur als chronisch progressiv bezeichnete Multiple Sklerose (CPMS) aufgeteilt wurde in RRMS und SPMS (Bryant 2001, S.578 [64]). Daneben werden Begriffe wie „exacerbating relapsing multiple sclerosis“ (ERMS), „progressive Multiple Sklerose“ (PMS), „relapsing multiple sclerosis“ (RMS) oder „relapsing-progressive multiple sclerosis“ (RPMS) verwendet. Um zukünftige Studien in höherem Maße vergleichbar zu machen, sollte als ein erster Schritt versucht werden, die Nomenklatur noch mehr aufeinander abzustimmen. Außerdem wäre für eine Verbesserung der Übereinstimmung von Studien denkbar, Standardeinschlusskriterien zu

definieren, die in jeder Studie zumindest in einer gut beschriebenen Subgruppenanalyse verwendet werden sollten. Vor dem Hintergrund der genannten Probleme werden im Folgenden nochmals die wichtigsten Informationen zu den Therapieformen kurz diskutiert.

Für **Beta-Interferone** konnten mehrere Studien einen Behandlungserfolg in der MS-Therapie nachweisen. Sowohl in der Gruppe der RRMS- als auch der SPMS-Patienten waren Beta-Interferone wirksam, wobei für SPMS-Patienten keine Daten zu IFN- β 1a existieren und die Validität der Studie, die IFN- β 1b bei RRMS untersuchte (The IFNB Multiple Sclerosis Study Group 1993 [150] und 1995 [149]), angezweifelt wird (Clegg 2000, S.13 [67]). Daher können für IFN- β 1a bei SPMS und für IFN- β 1b bei RRMS keine oder nur ungenügende Aussagen gemacht werden. In der MR-Untersuchung wurde für alle drei Interferone eine Abnahme der Läsionenzahl, Verminderung der Zahl der aktiven Scans oder Reduktion des Auftretens neuer Läsionen beobachtet. Gold et al. 2000 [84] ergänzen die Aussagen zu Beta-Interferonen bei schubförmiger MS: „Nach 2 Jahren Behandlung schnitten die behandelten Gruppen in Hinblick auf Schubfrequenz ca. 30% besser als die Placebogruppe ab. ... Die kernspintomographisch nachgewiesene Krankheitsaktivität nahm mit beiden Substanzen deutlich ab (zwischen 50 und 80%). Für das Ausmaß der Behinderung zeichnete sich schon in der kurzen Zeit für alle IFN- β -Präparate zumindest ein Trend zugunsten der Behandlungsgruppen ab“ (Gold 2000, S.69 [84]). Für sekundär progressive MS stellen Gold et al. 2000 fest, dass sich „die Ergebnisse ... erstaunlich gut mit der IFN- β 1b-Studie bei schubförmiger MS (European Study Group on Interferon β 1b 1998 [72]) ...“ (Gold 2000, S.69 [84]) vergleichen lassen. Die Autoren kommen zu dem Schluss, „... daß auch Patienten mit sekundär chronisch-progressiver MS bei Nachweis entzündlicher Krankheitsaktivität ... erfolgreich mit Interferon- β -Präparaten behandelt werden können“ (Gold 2000, S.69f. [84]).

Zu **Azathioprin** berichten alle verwendeten Studien über nicht-signifikante Verzögerungen der Krankheitsprogression (Clegg 2000, S.9-11 [67]). In dem sieben Studien umfassenden Review von Yudkin et al. 1991 wird eine Reduktion der Schubrate um ein Drittel oder mehr (Clegg 2000, S.10 [67]) errechnet. Allerdings fiel „... dieser (Effekt) geringer ... (aus) als die für die verschiedenen Interferone beschriebene schubreduzierende Wirkung“ (Gold 2000, S.76 [84]). Aufgrund des ungünstigen Nebenwirkungsprofils kann Azathioprin nur mit Einschränkungen als geeignet für die Therapie der MS gelten. Denn „... unpleasant side effects are common and some patients are unable to tolerate them“ (Bryant 2001, S.578 [64]). Azathioprin hat jedoch den Vorteil, dass es „... die einzig verfügbare orale Substanz mit langjähriger Anwendungserfahrung...“ (Wötzel 2000, S.42-44 [158]) ist. Derzeit wird das Medikament besonders bei Patienten mit häufigen Exazerbationen oder mit einem mittleren bis schweren Verlauf der MS verwendet (Clegg 2000, S.5 [67]).

Bei einer Therapie mittels **Lymphknotenbestrahlung** lässt sich für die in Cook 1997 [3] angegebenen Zielkriterien EDSS-Score, Exazerbationsrate und Verbesserung von MRT-Ergebnissen im Vergleich zu der reinen Steroidgabe eine gute Wirksamkeit der TLI mit insgesamt 1980 cGY in Kombination mit 30mg/d Kortikosteroiden feststellen. Die Validität der genannten Resultate wird durch eine teilweise Aufhebung der Verblindung zwar etwas getrübt, dennoch sind die Ergebnisse vielversprechend und sollten in zukünftigen Studien verifiziert werden. Daher muss vorerst auch Gold 2000 [84] widersprochen werden, der behauptet, dass Cook 1997 [3] „keine überzeugende(n) Beweise für die Wirksamkeit der Therapie liefern“ (Gold 2000, S.81 [84]) konnte.

Als weitere Therapieoption ist **Glatirameracetat** zu nennen. Dieses bietet eine Behandlungswirkung auf die Rate der Exazerbationen und zum Teil auch auf das Fortschreiten der Erkrankung. Da die Validität der in Clegg 2000 [67] zugrundeliegenden Studien eingeschränkt ist (Clegg 2000, S.21 [67]), sollten weitere Hochqualitätsstudien folgen, um die therapeutische Wirkung von Glatirameracetat zu überprüfen. Eine neue Studie auf diesem Gebiet, die in Gold 2000 [84] erwähnt wurde, aber noch nicht diskutiert werden konnte, führte die Arbeitsgruppe von Filippi/Comi 1999 durch. Darin konnte „... ein statistisch hochsignifikanter Rückgang der floriden, Gadolinium-aufnehmenden Läsionen um 35% ...“ (Gold 2000, S.72 [84]) gezeigt werden. Jedoch werden auch damit keine weiteren Aussagen über klinische Parameter wie zum Beispiel die Krankheitsprogression gemacht.

Zur Gabe von **Mitoxantron** ist festzustellen, dass dieser Stoff vielversprechende Resultate im klinischen Bereich der Zielkriterien liefert, bei den MRT-Ergebnissen in Millefiorini 1997 jedoch noch keine signifikante Wirkung beweisen konnte (Millefiorini 1997, Tabelle 4, S.156 [19]; Clegg 2000, S.25-28 [67]). Nach Gold 2000 existiert mittlerweile eine „... Klasse-I-Evidenz-Studie [=Hartung 1999 [196]] bei Patienten mit schwerem schubförmigen oder schubförmig-progredientem Krankheitsverlauf ...“ (Gold 2000, S.78 [84]), die auch einen dosisabhängigen „... Effekt auf die kernspintomographischen Verlaufsparemeter wie Gadolinium anreichernde Herde und T2-Gesamtläsionslast (feststellt) Somit steht mit dem Mitoxantron ein alternatives wirksames therapeutisches Prinzip zur Verfügung, welches bei Versagen von Basistherapeutika zur immunprophylaktischen Behandlung der Multiplen Sklerose eingesetzt werden kann“ (Gold 2000, S.78 [84]).

Ein weiterer experimenteller Ansatz der MS-Therapie mit teilweise positiven Ergebnissen ist die intravenöse monoklonale Antikörpertherapie mit **Immunglobulin G**. Aus den Studienresultaten leitet sich folgende Aussage ab: Es kommt unter IgG-Therapie zu einer positiven Beeinflussung der Häufigkeit der Exazerbationen, wobei eher kein signifikanter Einfluss auf die Krankheitsprogression besteht. Außerdem hat IgG „... a wide range of side effects“ (Bryant 2001, S.578 [64]). Ein Erfolg der IgG-Therapie sollte aber noch in weiteren größeren kontrollierten randomisierten Studien mit längeren Beobachtungszeit-

räumen evaluiert werden. Daher muss die IgG-Behandlung derzeit noch zur oben beschriebenen Gruppe 2 der MS-Therapieoptionen gezählt werden.

Cladribin ist ebenfalls zur zweiten Gruppe der Stoffe zu zählen: Eine Wirksamkeit konnte nur teilweise nachgewiesen werden. Romine et al. 1999 [204] fanden eine signifikante Reduktion der Häufigkeit in Kombination mit der Schwere der Schübe (Clegg 2000, S.17 [67]). Deutliche Unterschiede im EDSS- oder SNRS-Score konnte jedoch keine der drei von Clegg und Kollegen verwendeten Studien beweisen (Clegg 2000, S.6, 16f. und 59-62 [67]). Weder Gold 2000 [84] noch Wötzel 2000 [158] machen Angaben zu Cladribin als Therapieoption bei MS. Wegen widersprüchlicher Ergebnisse und möglicher Toxizität kann dieses Medikament für die Therapie der MS noch nicht empfohlen werden.

Auch **Methotrexat** sollte für die MS-Behandlung derzeit im Allgemeinen nicht eingesetzt werden. Laut Gold und Kollegen stellt es lediglich bei SPMS eine Therapiealternative dar, „... wobei allerdings der günstige Effekt nur auf die Erhaltung der Funktion der oberen Extremitäten ausgeprägt ist“ (Gold 2000, S.79f. [84]). Es lässt sich feststellen, dass entsprechend der Resultate aus Currier 1993 [193] Methotrexat keinen Einfluss auf die Rate der Exazerbationen und nur eine nicht-signifikante Wirkung auf die positive Entwicklung des EDSS-Scores hat. Lediglich für das Konzept des Behandlungsmisserfolgs, an dem verschiedene Bewertungsmaßstäbe beteiligt sind, kann ein signifikanter Therapieerfolg für Methotrexat im Vergleich mit der Placebogruppe nachgewiesen werden (Clegg 2000, S.25f. [67]).

Glukokortikoide konnten ihre Wirkung auf das schnellere Abklingen von Schüben zwar beweisen, der Erfolg bei Langzeittherapie bleibt jedoch ungewiss (Gold 2000, S.63f. [84]). Aufgrund mangelnder Validität einiger Therapiestudien mit Prednisolon und geringer Aussagekraft (Gold 2000, S.63 [84]) sind neue, größere und bessere Therapiestudien zu fordern, aus denen eindeutige Ergebnisse zu erhoffen wären.

Außer den bisher genannten Therapieoptionen der MS, die einen mindestens teilweise vorhandenen Behandlungserfolg deutlich machen, existieren eine Reihe von Untersuchungen, deren experimentelle Behandlungsgruppe kein besseres Outcome als die Placebogruppe aufweisen kann. So ist **Lenercept** aufgrund der schlechten Ergebnisse in der Therapie der MS wohl kontraindiziert (The Lenercept Multiple Sclerosis Study Group 1999, S.463f. [41])

Auch eine Therapie mit **Sulfasalazin** scheint aufgrund der Ergebnisse von Noseworthy und Kollegen nicht sinnvoll: Dieses Therapiekonzept hatte nach ein und zwei Jahren vielversprechende Resultate geliefert, nach drei Jahren lag die Progressionsrate der Sulfasalazin-Gruppe jedoch über derjenigen der Placebogruppe (Noseworthy 1998, S.1346 [27]).

Zu **AG284** ist festzustellen, dass derzeit noch keine ausreichende Evidence besteht, um es in der Therapie der MS einsetzen zu können. Die bis zum Jahr 2000 einzige existierende randomisierte kontrollierte Studie zu diesem Thema hatte nicht die Power, um Behandlungseffekte erkennen zu können (Goodkin 2000, S.1419 [10]).

Aus den Resultaten von Tubridy 1999 kann abgeleitet werden, dass bis zu dem Zeitpunkt, an dem neue Erkenntnisse aus größeren Studien mit längerem FU vorliegen, **Antegren** wegen fehlender Behandlungseffekte für die Therapie der MS vorerst als nicht geeignet erscheint.

Cyclosporin A zeigte zwar eine dem Azathioprin vergleichbare Wirkung (Gold 2000, S.79 [84]; Kappos 1988 [198]), dennoch sollte es für die Therapie der MS nur in Ausnahmefällen eingesetzt werden, da in den Nebenwirkungen „... eine deutliche Limitierung dieser Behandlungsform besteht“ (Gold 2000, S.79 [84]).

Von einer MS-Behandlung mit **Cyclophosphamid** muss aufgrund der schlechten Studienlage, der nicht-signifikanten Resultate und der Nebenwirkungen im Allgemeinen ebenfalls abgeraten werden. Dennoch setzen Gold und Kollegen diese Therapie „... bei chronisch-progredient verlaufender MS ... ein, wenn die Erkrankung ein rasches Vorschreiten zeigt und auf andere, besser verträgliche und weniger toxische Alternativen nicht anspricht“ (Gold 2000, S.80 [84]). Cyclophosphamid sollte also als „ultima ratio“ nur bei Versagen aller anderen möglichen Alternativen verwendet werden.

Aus den hier vorgestellten Therapieoptionen soll nun als Schritt 3 der EBM-Anwendung für den anfangs im Szenario der Einleitung vorgestellten fiktiven Patienten die passende Behandlungsform diskutiert werden. Die Einschlusskriterien betreffend, kommen mit Ausnahme der Studie zur Lymphknotenbestrahlung (Cook 1997 [3]), in der erst Patienten ab einem EDSS von 3,5 teilnahmen, alle der für das Review verwendeten Studien in Betracht. Wie oben beschrieben, können jedoch nur drei Stoffe derzeit eine Wirksamkeit eindeutig nachweisen, unter anderem eine Lymphknotenbestrahlung (TLI) in Kombination mit Steroiden, die für den Patienten aus dem Szenario wegen unterschiedlicher Einschlusskriterien und noch nicht erfolgter Bestätigung der Ergebnisse eher nicht in Frage kommt. Aufgrund des ungünstigen Nebenwirkungsprofils sollte auch Azathioprin erst als zweite Wahl in der Therapie des Patienten eingesetzt werden. Für IFN- β 1b können bei RRMS wegen mangelnder Validität der entsprechenden Studie (Clegg 2000, S.13 [67]) und nicht übereinstimmender Kriterien mit Jacobs und Kollegen, NEJM 2000, die Patienten mit beginnender MS untersuchten, keine oder nur ungenügende Aussagen gemacht werden. IFN- β 1a hingegen konnte seine Wirksamkeit bei RRMS-Patienten beweisen, weshalb der Patient entweder mit **Avonex, 30 μ g 1x/ Woche i.m.**, oder **Rebif, 44 μ g 3x/ Woche s.c.**, behandelt werden sollte. Je nach seiner Präferenz der i.m.- oder s.c.-

Injektion und der Häufigkeit der Einnahme würde dann entschieden, welches Medikament verabreicht wird.

4.5. Stellenwert der Magnetresonanztomographie bei Multipler Sklerose

Derzeit wird die MRT in Pilot-/Phase1-Studien als primäres Zielkriterien angewandt, da die hohe Sensitivität von monatlichen T₂-gewichteten und Gd-verstärkten T₁- MR-Scans bei der Entdeckung von klinisch asymptomatischen Läsionen es möglich macht, in kürzeren Studiendauern eventuelle Behandlungserfolge besser sichtbar zu machen als es mit klinischen Outcomes der Fall wäre (Evans 1997, S.126 [73]). Quantitative Untersuchung von Läsionen ist verlässlich und reproduzierbar in der Überwachung von Behandlungseffekten (Filippi 1995, S.1609 [78]). Da die Korrelationen zwischen dem Auftreten von Läsionen und Klinik der MS nur in Teilbereichen gelten und der Zusammenhang zwischen Läsionenlast und späterer Krankheitsentwicklung noch nicht sicher bestätigt ist, kann die MRT für Therapiestudien späterer Phasen der Medikamentenzulassung bisher aber nur als sekundäres Zielkriterium empfohlen werden (Miller 1996, S.9 [104]; Evans 1997, S.126 [73]).

Laut Filippi und Kollegen besteht in vielen Bereichen der MRT Verbesserungsbedarf, so zum Beispiel bei der Korrelation zwischen der Ausdehnung von Auffälligkeiten im MRT und der klinischen Entwicklung der Erkrankung (Filippi 1995, S.1609 [78]). Dies könnte einerseits erreicht werden durch eine Veränderung der klinischen Beurteilungsskalen und andererseits durch neue MR-Techniken, die spezifischere Ergebnisse bezüglich der Pathologie liefern (Filippi 1995, S.1609 [78]). Mittlerweile existieren eine Reihe modifizierter MR-Anwendungen. Zu den T₂-gewichteten Möglichkeiten gehören „Fast-spin-echo“- Bilder (FSE) mit dem Vorteil, dünne aneinander angrenzende Schichten mit hohen Kontrastunterschieden erreichen zu können (Miller 1998, S.5 [103]) und „FLAIR“ (fluid attenuated inversion recovery), das auch Läsionen an der Grenze zwischen Gehirn und Liquor entdecken kann (Miller 1998, S.5f. [103]). Mit Gadolinium als Kontrastmittel können eine Störung der Blut-Hirn-Schranke sowie inflammatorische Prozesse sehr gut sichtbar gemacht werden (Miller 1998, S.6 [103]; Kappos 1999, S.967 [95]). Damit bieten Gd-T₁-Bilder eine äußerst sensitive Möglichkeit, Veränderungen in diesem Bereich früh erkennen zu können. Der Nachteil dieser Art der MRT ist eine geringe Spezifität (Miller 1998, S.6 [103]). Außerdem werden Studien mit längerer Beobachtungszeit als bisher benötigt, um zu entscheiden, ob die aus Studien mit Gd-T₁-MRT gewonnenen prognostischen Daten als Ergänzung zu T₂-Bildern nützlich sind (Miller 1996, S.8 [104]). Auch „Magnetization Transfer“ (MT) wird bei MS eingesetzt. Die Vorteile liegen in einer hohen Reproduzierbarkeit

innerhalb eines Zentrums und einer hohen Sensitivität. Als Nachteil gilt einerseits die Abhängigkeit von Bewegungen des Patienten und andererseits, dass es sich bei MRT um keine absolute Messmethode handelt, das heißt, die Durchführung erfordert bestimmte, frei wählbare Voreinstellungen (Miller 1998, S.7 [103]).

Generell gilt, dass es immer noch Schwierigkeiten bei der Korrelation zwischen T₂-MRT und klinischem Verlauf bei MS-Patienten gibt. Miller und Kollegen nennen dafür fünf Ursachen (Miller 1996, S.8 [104]):

1. Viele Läsionen führen nicht direkt zu einer Behinderung der Bewegungsfähigkeit. Diese spielt aber bei der Evaluation des EDSS-Scores eine besonders wichtige Rolle.
2. Die Methoden sowohl zur Bestimmung der Läsionenlast als auch der Behinderung sind nicht exakt.
3. Frühere Studien hatten zu kurze Beobachtungszeiten, um eine Änderung der Behinderung feststellen zu können.
4. Unsichtbare Veränderungen in der weißen Substanz führen zu Behinderungen. Zwar sind viele dieser Veränderungen mittlerweile mit den oben genannten modernen MR-Techniken sichtbar zu machen, dennoch ist das Ausmaß, in dem die dann erkannten Läsionen mit den Bewegungseinschränkungen zusammenhängen, bisher unklar.
5. Im herkömmlichen MRT gleich aussehende Läsionen enthalten unterschiedliche pathophysiologische Substrate.

Es muss also die Bestimmung von Zahl und Größe der Läsionen sowie die Korrelation der Ergebnisse mit dem Grad der Behinderung verbessert werden, um eine höhere Objektivität zu erreichen. Für eine Verbesserung der Korrelation kommen die oben genannten alternativen MR-Techniken in Frage. Die Beurteilung von Läsionenzahl und -ausdehnung geschieht am besten mit einer Computer-unterstützten Technik. Allerdings bestehen auch hier Probleme. So sind Läsionen bei MS eher unscharf begrenzt und variieren stark in ihrem Erscheinungsbild. Läsionen können klein aber zahlreich sein und die Qualität der Bilder leidet unter Artefakten wie Bewegungen des Patienten, Eigenarten des MR-Geräts, Blut- oder Liquorfluss. Es gibt keine Technik, mit der man die wahre Läsionengröße- und zahl bestimmen könnte, da es derzeit keinen Goldstandard gibt (Miller 1998, S.9 [103]). Dennoch sollte die menschliche Bewertung soweit wie möglich vermieden werden (Filippi 1995, S.1609 [78]), da eine Computerauswertung generell als objektiver zu betrachten ist. Schließlich müssen auf der klinischen Seite für eine Verbesserung der Korrelation die Beurteilungsskalen verändert werden (Filippi 1995, S.1609 [78]), um mehr auf die wahre Pathologie einzugehen und die Überbewertung der Gehfunktion wie im EDSS-Score zurückzudrängen (Kappos 1999, S.967 [95]).

5. Zusammenfassung

Ziel der vorliegenden Arbeit war es, aus den Ergebnissen der Bewertung eingeschlossener Multiple-Sklerose-Therapiestudien Verbesserungsvorschläge für Durchführung und Berichterstattung zukünftiger Studien zu erarbeiten. Dabei konnte der Nutzen der Evidence-based Medicine für eine wissenschaftlich gestützte objektive Medizin belegt und die Wichtigkeit der Forderungen des CONSORT-Statements als Voraussetzung für eine objektive Validitätsbewertung nach EBM-Kriterien bewiesen werden. Den Autoren zukünftiger klinischer Therapiestudien ist die Umsetzung der CONSORT-Forderungen, am besten berichtet in einer Liste, in der das Vorgehen zur Erfüllung aller 22 CONSORT-Punkte beschrieben wird (siehe Abbildung 3 auf Seite 19 dieser Arbeit) dringend anzuraten. Nur so kann den Ergebnissen Aussagekraft verliehen werden und der kritische Leser eindeutige Bewertungskriterien erhalten.

Die vorliegende Arbeit bietet eine vollständige Beurteilung aller Forderungen des CONSORT-Statements kombiniert mit einer Auswertung der Validitäts-Bewertungsergebnisse nach dem Sackettschen EBM-Schema zusammen mit dem objektivierten Beurteilungsscore nach Hassler. Dabei wurde deutlich, dass selbst die besonders wichtigen EBM-Bereiche Randomisierung und Verblindung, denen nachgewiesen werden konnte, dass bei Nichtbeachtung der Erfolg der untersuchten Therapie überschätzt wird (Schulz 1995, S.410 [140]), immer noch viel zu häufig nicht oder nur teilweise erfüllt werden. Eine verbesserte Berichterstattung entsprechend der CONSORT-Forderungen Nummer 8 und 9 für die Randomisierung (Altman 2001, S.671ff. [48]) und Nummer 11 für die Verblindung (Altman 2001, S.674ff. [48]) ist in zukünftigen Therapiestudien als äußerst wichtig zu erachten.

Die Fortentwicklung des Sackettschen EBM-Schemas in der vorliegenden Arbeit zu dem objektivierten Bewertungssystem nach Hassler bietet ein Instrument zur noch besseren und reproduzierbareren Beurteilung der Validität von Therapiestudien. Wo immer möglich, wurden den Fragen nach Sackett 1999 [134] Bewertungspunkte gegenübergestellt, um eine direkte Vergleichbarkeit der Validität von Therapiestudien zu ermöglichen. Damit kann die Beurteilung in Zukunft noch objektiver als bisher erfolgen.

In dem an die EBM-Auswertung anschließenden Review der Therapieoptionen bei MS zeigt sich der Nutzen der Beta-Interferone und des Azathioprins. Auch wenn eine modifizierte totale Lymphknotenbestrahlung zusammen mit Kortikosteroiden ihre Wirksamkeit bereits in einer Studie nachweisen konnte, kann diese Therapie derzeit noch nicht empfohlen werden, da sie in weiteren Studien verifiziert werden muss. Mitoxantron und Glatirameracetat, das in Deutschland bereits zugelassen ist, scheinen ebenfalls vielverspre-

chend, müssen jedoch in größeren validen Studien überprüft werden. Cladribin, Methotrexat, Immunglobulin G und Glukokortikoide konnten ihre Wirksamkeit in der Langzeittherapie noch nicht beweisen, und Lenercept, Sulfasalazin, AG284, Antegren, Cyclophosphamid sowie Cyclosporin A sollten zur Therapie der MS entweder wegen fehlender Wirksamkeit oder zu hoher Toxizität nicht eingesetzt werden.

Für den Einsatz der MRT bei Multipler Sklerose gilt, dass sowohl der Zusammenhang zwischen Läsionenzahl und späterer Entwicklung der Erkrankung als auch die Korrelation mit dem EDSS-Score bei Patienten mit bereits bestehender MS nicht sicher bestätigt werden kann. Bei Patienten mit einem ersten demyelinisierenden Symptom findet sich jedoch in den meisten Studien ein statistisch relevanter Zusammenhang zwischen der Läsionenzahl und -lage mit der späteren Entwicklung klinisch definitiver MS. Damit ist eine Vorhersage bei ersten Symptomen einer MS möglich, während die Diagnostik bei bereits bestehender Multipler Sklerose weiterhin primär auf klinischen Parametern bestehen sollte und die MRT dabei nur Unterstützung sein kann. Auch wenn derzeit noch manche Probleme bestehen, erhöhen neue MRT-Techniken, verbesserte Auswertungsmöglichkeiten per Computer und Scans von Hirnstamm und Rückenmark die Aussagekraft.

Douglas Altman, ein äußerst wichtiger Verfechter der Evidence-based Medicine schreibt in einem Artikel: „We need less research, better research, and research done for the right reasons. Abandoning using the number of publications as a measure of ability would be a start.“ (Altman 1994, S.284 [52]) Die EBM ist heute schon nicht mehr aus der Medizin wegzudenken, doch trotz der Forderung nach wissenschaftlich gestützter, objektiver Medizin werden einige der EBM-Kriterien immer noch viel zu selten beachtet. Ein tieferes Vordringen in die Köpfe klinischer Forscher ist nötig, um auch die letzten Zweifel an der Wirksamkeit neuer, in klinischen Studien überprüfter Therapieoptionen auszuräumen.

6. Literaturverzeichnis

6.1. Primärliteratur: Bewertete Studien

1. Brunetti A, Tedeschi G, Di Costanzo A, Covelli EM, Aloj L, Bonavita S, Ciarmiello A, Alfano B, Salvatore M: White matter lesion detection in multiple sclerosis: improved interobserver concordance with multispectral MRI display. *J Neurol* 1997; 244(9):586-90
2. Coles AJ, Wing MG, Molyneux P, Paolillo A, Davie CM, Hale G, Miller D, Waldmann H, Compston A: Monoclonal antibody treatment exposes three mechanisms underlying the clinical course of multiple sclerosis. *Ann Neurol* 1999; 46(3):296-304
3. Cook SD, Devereux C, Troiano R, Wolansky L, Guarnaccia J, Haffty B, Bansil S, Goldstein J, Sheffet A, Zito G, Jotkowitz A, Boos J, Dowling P, Rohowsky-Kochan C, Volmer T: Modified total lymphoid irradiation and low dose corticosteroids in progressive multiple sclerosis. *J Neurol Sci* 1997; 152(2):172-181
4. Edan G, Miller D, Clanet M, Confavreux C, Lyon-Caen O, Lubetzki C, Brochet B, Berry I, Rolland Y, Froment JC, Cabanis E, Iba-Zizen MT, Gandon JM, Lai HM, Moseley I, Sabouraud O: Therapeutic effect of mitoxantrone combined with methylprednisolone in multiple sclerosis: a randomised multicentre study of active disease using MRI and clinical criteria. *J Neurol Neurosurg Psychiatry* 1997; 62(2):112-118
5. Edwards S, Zvartau M, Clarke H, Irving W, Blumhardt LD: Clinical relapses and disease activity on magnetic resonance imaging associated with viral upper respiratory tract infections in multiple sclerosis. *J Neurol Neurosurg Psychiatry* 1998; 64(6):736-41
6. Filippi M, Mastrorlando G, Bastianello S, Rocca MA, Rovaris M, Gasperini C, Pozzilli C, Comi G: A longitudinal brain MRI study comparing the sensitivities of the conventional and a newer approach for detecting active lesions in multiple sclerosis *J Neurol Sci* 1998; 159(1):94-101
7. Filippi M, Rovaris M, Iannucci G, Mennea S, Sormani MP, Comi G: Whole brain volume changes in patients with progressive MS treated with cladribine. *Neurology* 2000; 55(11):1714-1718
8. Filippi M, Rovaris M, Rice GP, Sormani MP, Iannucci G, Giacomotti L, Comi G: The effect of cladribine on T(1) 'black hole' changes in progressive MS. *J Neurol Sci* 2000; 176(1):42-44
9. Fisher E, Rudick RA, Cutter G, Baier M, Miller D, Weinstock-Guttman B, Mass MK, Dougherty DS, Simonian NA: Relationship between brain atrophy and disability: an 8-year follow-up study of multiple sclerosis patients. *Mult Scler* 2000; 6(6):373-7
10. Goodkin DE, Shulman M, Winkelhake J, Waubant E, Andersson P, Stewart T, Nelson S, Fischbein N, Coyle PK, Frohman E, Jacobs L, Holcenberg J, Lee M, Mocchi S: A phase I trial of solubilized DR2:MBP84-102 (AG284) in multiple sclerosis. *Neurology* 2000; 54(7):1414-1420
11. Hommes OR, Maas-Enriquez M: ESIMS—an ongoing clinical trial in secondary progressive multiple sclerosis. *Mult Scler* 2000; 6 Suppl 2:S27-32
12. Jacobs L, Rudick R, Simon J: Extended observations on MS patients treated with IM interferon-beta1a (Avonex): implications for modern MS trials and therapeutics. *J Neuroimmunol* 2000; 107(2):167-173
13. Jacobs LD, Beck RW, Simon JH, Kinkel RP, Brownscheidle CM, Murray TJ, Simonian NA, Slasor PJ, Sandrock AW: Intramuscular interferon beta-1a therapy initiated during a first demyelinating event in multiple sclerosis, CHAMPS Study Group. *N Engl J Med* 2000; 343(13):898-904
14. Kappos L, Comi G, Panitch H, Oger J, Antel J, Conlon P, Steinman L: Induction of a non-encephalitogenic type 2 T helper-cell autoimmune response in multiple sclerosis after administration of an altered peptide ligand in a placebo-controlled, randomized phase II

- trial, The Altered Peptide Ligand in Relapsing MS Study Group. *Nat Med* 2000;6(10):1176-1182
15. Kastrukoff LF, Morgan NG, Zecchini D, White R, Petkau AJ, Satoh J, Paty DW: Natural killer cells in relapsing-remitting MS: effect of treatment with interferon beta-1B. *Neurology* 1999; 52(2):351-359
 16. Koch-Henriksen N, Sørensen PS: The Danish National Project of interferon-beta treatment in relapsing-remitting multiple sclerosis, The Danish Multiple Sclerosis Group. *Mult Scler* 2000; 6(3):172-175
 17. Koudriavtseva T, Pozzilli C, Fiorelli M, Gasperini C, Bagnato F, Galgani S, Frontoni M, Ciccarelli O, Bastianello S: Determinants of Gd-enhanced MRI response to IFN-beta-1a treatment in relapsing-remitting multiple sclerosis. *Mult Scler* 1998; 4(5):403-407
 18. Li DK, Paty DW: Magnetic resonance imaging results of the PRISMS trial: study of subcutan interferon-beta1a in relapsing-remitting multiple sclerosis; Prevention. *Ann Neurol* 1999; 46(2):197-206
 19. Millefiorini E, Gasperini C, Pozzilli C, D'Andrea F, Bastianello S, Trojano M, Morino S, Morra VB, Bozzao A, Calo' A, Bernini ML, Gambi D, Prencipe M: Randomized placebo-controlled trial of mitoxantrone in relapsing-remitting multiple sclerosis: 24-month clinical and MRI outcome. *J Neurol* 1997; 244(3):153-159
 20. Miller DH, Molyneux PD, Barker GJ, MacManus DG, Moseley IF, Wagner K: Effect of interferon-beta1b on magnetic resonance imaging outcomes in secondary progressive multiple sclerosis: results of a European multicenter, randomized, double-blind, placebo-controlled trial. *Ann Neurol* 1999; 46(6):850-859
 21. Molyneux PD, Brex PA, Fogg C, Lewis S, Middleditch C, Barkhof F, Sormani MP, Filippi M, Miller DH: The precision of T1 hypointense lesion volume quantification in multiple sclerosis treatment trials: a multicenter study. *Mult Scler* 2000; 6(4):237-40
 22. Molyneux PD, Kappos L, Polman C, Pozzilli C, Barkhof F, Filippi M, Yousry T, Hahn D, Wagner K, Ghazi M, Beckmann K, Dahlke F, Losseff N, Barker GJ, Thompson AJ, Miller DH: The effect of interferon beta-1b treatment on MRI measures of cerebral atrophy in secondary progressive multiple sclerosis, European Study Group on Interferon beta-1b in secondary progressive multiple sclerosis. *Brain* 2000; 123 (Pt 11):2256-2263
 23. Molyneux PD, Tofts PS, Fletcher A, Gunn B, Robinson P, Gallagher H, Moseley IF, Barker GJ, Miller DH: Precision and reliability for measurement of change in MRI lesion volume in multiple sclerosis: a comparison of two computer assisted techniques. *J Neurol Neurosurg Psychiatry* 1998; 65(1):42-47
 24. Mycko MP, Kwinkowski M, Tronczynska E, Szymanska B, Selmaj KW: Multiple sclerosis: the increased frequency of the ICAM-1 exon 6 gene point mutation genetic type K469. *Ann Neurol* 1998; 44(1):70-75
 25. Myhr KM, Riise T, Green Lilleas FE, Beiske TG, Celius EG, Edland A, Jensen D, Larsen JP, Nilsen R, Nortvedt MW, Smievoll AI, Vedeler C, Nyland HI: Interferon-alpha2a reduces MRI disease activity in relapsing-remitting multiple sclerosis, Norwegian Study Group on Interferon-alpha in Multiple Sclerosis. *Neurology* 1999; 52(5):1049-1056
 26. Nortvedt MW, Riise T, Myhr KM, Nyland HI, Hanestad BR: Type I interferons and the quality of life of multiple sclerosis patients. *Mult Scler* 1999; 5(5):317-322
 27. Noseworthy JH, O'Brien P, Erickson BJ, Lee D, Sneve D, Ebers GC, Rice GP, Auty A, Hader WJ, Kirk A, Duquette P, Carter J, Francis G, Metz L, Shuster E: The Mayo Clinic-Canadian Cooperative trial of sulfasalazine in active multiple sclerosis. *Neurology* 1998; 51(5):1342-1352
 28. Oliveri RL, Valentino P, Russo C, Sibilgia G, Aguglia U, Bono F, Fera F, Gambardella A, Zappia M, Pardatscher K, Quattrone A: Randomized trial comparing two different high doses of methylprednisolone in MS: a clinical and MRI study. *Neurology* 1998; 50(6):1833-1836
 29. Optic Neuritis Study Group: The 5-year risk of MS after optic neuritis, Experience of the optic neuritis treatment trial. *Neurology* 1997; 49(5):1404-1413

30. Paolillo A, Bastianello S, Frontoni M, Gasperini C, Giugni E, Ciccarelli O, Luccichenti G, Cannoni S, Pozzilli C: Magnetic resonance imaging outcome of new enhancing lesions in relapsing-remitting multiple sclerosis patients treated with interferon beta1a. *J Neurol* 1999; 246(6):443-448
31. Patti F, L'Episcopo MR, Cataldi ML, Reggio A: Natural interferon-beta treatment of relapsing-remitting and secondary-progressive multiple sclerosis patients, A two-year study. *Acta Neurol Scand* 1999; 100(5):283-289
32. Rice GP, Filippi M, Comi G: Cladribine and progressive MS: clinical and MRI outcomes of a multicenter controlled trial, Cladribine MRI Study Group. *Neurology* 2000; 54(5):1145-1155
33. Rohowsky-Kochan C, Molinaro D, Devereux C, Troiano R, Bansil S, Zito G, Wolansky L, Jotkowitz A, Denny T, Oleske J, Dowling P, Cook SD: The effect of total lymphoid irradiation and low-dose steroids on T lymphocyte populations in multiple sclerosis: correlation with clinical and MRI status. *J Neurol Sci* 1997; 152(2):182-192
34. Rudick RA, Fisher E, Lee JC, Simon J, Jacobs L: Use of the brain parenchymal fraction to measure whole brain atrophy in relapsing-remitting MS, Multiple Sclerosis Collaborative Research Group. *Neurology* 1999; 53(8):1698-1704
35. Sailer M, Heinze HJ, Schoenfeld MA, Hauser U, Smid HG: Amantadine influences cognitive processing in patients with multiple sclerosis. *Pharmacopsychiatry* 2000; 33(1):28-37
36. Simon JH, Jacobs LD, Champion M, Wende K, Simonian N, Cookfair DL, Rudick RA, Herndon RM, Richert JR, Salazar AM, Alam JJ, Fischer JS, Goodkin DE, Granger CV, Lajaunie M, Martens-Davidson AL, Meyer M, Sheeder J: Magnetic resonance studies of intramuscular interferon beta-1a for relapsing multiple sclerosis, The Multiple Sclerosis Collaborative Research Group. *Ann Neurol* 1998; 43(1):79-87
37. Sørensen PS, Wanscher B, Jensen CV, Schreiber K, Blinkenberg M, Ravnborg M, Kirsmeier H, Larsen VA, Lee ML: Intravenous immunoglobulin G reduces MRI activity in relapsing multiple sclerosis. *Neurology* 1998; 50(5):1273-1281
38. Sørensen PS: The effect on MRI of gammaglobulin treatment in relapsing multiple sclerosis. *Mult Scler* 2000; 6 Suppl 2:S14-S17
39. Tan IL, Lycklama a Nijeholt GJ, Polman CH, Ader HJ, Barkhof F: Linomide in the treatment of multiple sclerosis: MRI results from prematurely terminated phase-III trials. *Mult Scler* 2000; 6(2):99-104
40. Teksam M, Tali T, Kocer B, Isik S: Qualitative and quantitative volumetric evaluation of the efficacy of intravenous immunoglobulin in multiple sclerosis: preliminary report. *Neuroradiology* 2000; 42(12):885-889
41. The Lenercept Multiple Sclerosis Study Group and The University of British Columbia MS/MRI Analysis Group: TNF neutralization in MS: results of a randomized, placebo-controlled multicenter study. *Neurology* 1999; 11;53(3):457-465
42. The Once Weekly Interferon for MS Study Group: Evidence of interferon beta-1a dose response in relapsing-remitting MS: the OWIMS Study. *Neurology* 1999; 53(4):679-686
43. Tubridy N, Behan PO, Capildeo R, Chaudhuri A, Forbes R, Hawkins CP, Hughes RA, Palace J, Sharrack B, Swingle R, Young C, Moseley IF, MacManus DG, Donoghue S, Miller DH: The effect of anti-alpha4 integrin antibody on brain lesion activity in MS, The UK Antegren Study Group. *Neurology* 1999; 53(3):466-472
44. van Oosten BW, Lai M, Hodgkinson S, Barkhof F, Miller DH, Moseley IF, Thompson AJ, Rudge P, McDougall A, McLeod JG, Ader HJ, Polman CH: Treatment of multiple sclerosis with the monoclonal anti-CD4 antibody cM-T412: results of a randomized, double-blind, placebo-controlled, MR-monitored phase II trial. *Neurology* 1997; 49(2):351-357
45. Wolinsky JS, Narayana PA, Noseworthy JH, Lublin FD, Whitaker JN, Linde A, Gjorstrup P, Sullivan HC: Linomide in relapsing and secondary progressive MS: part II: MRI results, MRI Analysis Center of the University of Texas-Houston, Health Science Center, and the North American Linomide Investigators. *Neurology* 2000; 54(9):1734-1741

46. Zhao GJ, Koopmans RA, Li DK, Bedell L, Paty DW: Effect of interferon beta-1b in MS: assessment of annual accumulation of PD/T2 activity on MRI, UBC MS/MRI Analysis Group and the MS Study Group. *Neurology* 2000; 54(1):200-206

6.2. Sekundärliteratur

47. Achiron A, Gabbay U, Gilad R, Hassin Baer S, Barak Y, Gornish M, Elizur A, Goldhammer Y, Sarova-Pinhas I: Intravenous immunoglobulin treatment in multiple sclerosis. Effect on relapses. *Neurology* 1998; 50:398-402
48. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gøtzsche PC, Lang T, for the CONSORT group: The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001; 134:663-694
49. Altman DG: Better reporting of randomized controlled trials: the CONSORT statement. *BMJ* 1996; 313:570-571
50. Altman DG, Doré CJ: Randomisation and baseline comparisons in clinical trials. *Lancet* 1990; 335:149-153
51. Altman DG: Randomisation – Essential for reducing bias. *BMJ* 1991; 302:1481-1482
52. Altman DG: The scandal of poor medical research. *BMJ* 1994; 308:283-284
53. Andersen O, Lyckee J, Tolleson PO, Svenningsson A, Runmarker B, Linde AS, Åström M, Gjørstrup P, Ekholm S: Linomide reduces the rate of active lesions in relapsing-remitting multiple sclerosis. *Neurology* 1996; 47:895-900
54. Anderson DW, Ellenberg JH, Leventhal CM, Reingold SC, Rodriguez M, Silberberg DH: Revised estimate of the prevalence of multiple sclerosis in the United States. *Ann Neurol* 1992; 31:333-336
55. Andersson PB, Waubant E, Gee L, Goodkin DE: Multiple sclerosis that is progressive from the time of onset. *Arch Neurol* 1999; 56:1138-1142
56. Arnold DL, Matthews PM: MRI in the diagnosis and management of multiple sclerosis. *Neurology* 2002; 58:S23-S31
57. ash: Therapie der Multiple Sklerose – Den Axonverlust aufhalten. *Medical Tribune* 2000; Suppl.49:20
58. Ausejo M, Saenz A, Moher D: [CONSORT: Ein Versuch, die Qualität der Publikation klinischer Studien zu verbessern] (Editorial). *Aten Primaria* 1998; 21:351-352
59. Bailer JC III, Mosteller F: Guidelines for statistical reporting in articles for medical journals: amplification and explanations. *Ann Intern Med* 1988; 108:266-273
60. Barkhof F, Filippi M, van Waesberghe JH, Molyneux P, Rovaris M, Lycklama à Nijeholt G, Tubridy N, Miller DH, Yousry TA, Radue EW, Adèr HJ: Improving interobserver variation in reporting gadolinium-enhanced MRI lesions in multiple sclerosis. *Neurology* 1997; 49:1682-1688
61. Beck RW, Cleary PA, Anderson MM, Keltner JL, Shults WT, Kaufman DI, Buckley EG, Corbett JJ, Kupersmith MJ, Miller NR, Savino PJ, Guy JR, Trobe JD, McCrary JA, Smith CH, Chrousos GA, Thompson HS, Katz BJ, Brodsky MC, Goodwin JA, Atwell CW and the Optic Neuritis Study Group: A randomised, controlled trial of corticosteroids in the treatment of acute optic neuritis. *N Engl J Med* 1992; 326:581-588
62. Beck RW, Cleary PA, Trobe JD, Kaufman DI, Kupersmith MJ, Paty DW, Brown CH and the Optic Neuritis Study Group: The effect of corticosteroids for acute optic neuritis on the subsequent development of multiple sclerosis. *N Engl J Med* 1993; 329:1764-1769
63. Begg CB, Cho MK, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D, Stroup DF: Improving the quality of reporting of randomised controlled trials: the CONSORT statement. *JAMA* 1996; 276:637-639

64. Bryant J, Clegg A, Milne R: Systematic review of immunomodulatory drugs for the treatment of people with multiple sclerosis: Is there good quality evidence on effectiveness and cost? *J Neurol Neurosurg Psychiatry* 2001; 70:574-579
65. Cella DF, Dineen K, Arnason B, Reder A, Webster MA, Karabatsos G, Chang C, Lloyd S, Mo F, Stewart J, Stefoski D: Validation of the functional assessment of multiple sclerosis quality of life instrument. *Neurology* 1996; 47:129-139
66. Clade H: Evidenzbasierte Patientenversorgung – Ein rationales Entscheidungsinstrument. *Deutsches Ärzteblatt* 2000, Jg. 97; 49:A3316-A3320
67. Clegg A, Bryant J, Milne R: Disease-modifying drugs for multiple sclerosis: a rapid and systematic review. *Health Technol Assess* 2000; 4(9)
68. Cochrane AL. Effectiveness and Efficiency. Random Reflections on Health Services. London: Nuffield Provincial Hospitals Trust, 1972.
69. DerSimonian R, Charette LJ, McPeck B, Mosteller F: Reporting on methods in clinical trials. *NEJM* 1982; 306(22):1332-1337
70. Dianzani F, Antonelli G: Physiological mechanism of production and action of interferons in response to viral infection. *Adv Exp Med Biol* 1989; 257:47-60
71. Egger M, Juni P, Bartlett C; CONSORT Group (Consolidated Standards of Reporting of Trials: Value of flow diagrams in reports of randomized controlled trials. *JAMA* 2001; 285:1996-1999
72. European Study Group on Interferon β 1b in Secondary Progressive Multiple Sclerosis: Placebo-controlled multicentre randomised trial of interferon β -1b in treatment of secondary progressive multiple sclerosis. *Lancet* 1998; 352:1491-1497
73. Evans AC, Frank JA, Antel J, Miller DH: The role of MRI in clinical trials of multiple sclerosis: comparison of image processing techniques. *Ann Neurol* 1997; 41:125-132
74. Evidence-Based Medicine Working Group: User's Guides to Evidence-based Medicine, *JAMA* 1992; 268(17):2420-2425
75. Fazekas F, Deisenhammer F, Strasser-Fuchs S, Nahler G, Mamoli B, for the Austrian Immunoglobulin in Multiple Sclerosis Study Group: Randomised placebo-controlled trial of monthly intravenous immunoglobulin therapy in relapsing-remitting Multiple Sclerosis. *Lancet* 1997; 349:589-593
76. Filippi M, Gawne-Cain ML, Gasperini C, van Waesberghe JH, Grimaud J, Barkhof F, Sormani MP, Miller DH: Effect of training and different measurement strategies on the reproducibility of brain MRI lesion load measurement in multiple sclerosis. *Neurology* 1998; 50:238-244
77. Filippi M, Horsfield MA, Morrissey SP, MacManus DG, Rudge P, McDonald WI, Miller DH: Quantitative brain MRI lesion load predicts the course of clinically isolated syndromes suggestive of multiple sclerosis. *Neurology* 1994; 44:635-641
78. Filippi M, Horsfield MA, Tofts PS, Barkhof F, Thompson AJ, Miller DH: Quantitative assessment of MRI lesion load in monitoring the evolution of multiple sclerosis. *Brain* 1995; 118:1601-1612
79. Ford B, Tampieri D, Francis G: Long term follow up of acute transverse partial myelopathy. *Neurology* 1992; 42:250-252
80. Frederiksen JL, Larsson HBW, Olesen J, Henriksen O: Magnetic resonance imaging of the brain in patients with acute monosymptomatic optic neuritis. *Acta Neurol Scand* 1991; 83:343-350
81. Freemantle N, Mason JM, Haines A, Eccles MP: CONSORT: an important step toward evidence-based health care. *Ann Intern Med* 1997; 126:81-83
82. Gass A, Barker GJ, Kidd D, Thorpe JW, MacManus D, Brennan A, Tofts PS, Thompson AJ, McDonald WI, Miller DH: Correlation of magnetization transfer ratio with clinical disability in multiple sclerosis. *Ann Neurol* 1994; 36:62-67

83. Gass A, Radu EW, Filippi M, Kappos L: MRI follow-up examinations in multiple sclerosis: Guidelines for quality assurance. *RoFo Fortschritte auf dem Gebiet der Röntgenstrahlen und der Bildgebenden Verfahren* 1999; 170(6):581-586
84. Gold R, Rieckmann P: Pathogenese und Therapie der Multiplen Sklerose. 2.Auflage – Bremen: UNI-MED, 2000
85. Goodkin DE: EDSS reliability (letter). *Neurology* 1991; 41:332
86. Guyatt GH, Sackett DL, Cook DJ, for the Evidence-Based Medicine Working Group: User's Guides to the Medical Literature. II. How to Use an Article About Therapy or Prevention – A. Are the Results of the Study Valid? *JAMA* 1993; 270:2598-2601
87. Guyatt GH, Sackett DL, Cook DJ, for the Evidence-Based Medicine Working Group: User's Guides to the Medical Literature. II. How to Use an Article About Therapy or Prevention - B. What were the Results and Will They Help Me in Caring for My Patients? *JAMA* 1994; 271(1):59-63
88. Guyatt GH. Evidence-based medicine. *ACP J Club* 1991; 114(suppl 2): A-16
89. Hibberd PL, Haegert DG, Sadovnick AD: A Glossary of Terms Frequently Used in Epidemiology and Genetics. *Ann Neurol* 1994; 36(S2):S244-S247
90. Hollis S, Campbell F: What is meant by intention-to-treat analysis? Survey of published randomized controlled trials. *BMJ* 1999; 319:670-674
91. Huston P, Hoey J: CMAJ endorses the CONSORT statement. *Can Med Assoc J* 1996; 155:1277-1279
92. Jacobs LD, Cookfair DL, Rudick R, Herndon R, Richert J, Salazar AM, Fischer JS, Goodkin DE, Granger CV, Simon JH, Alam JJ, Bartoszak DM, Bourdette DN, Braiman J, Brownscheidle CM, Coats ME, Cohan SL, Dougherty DS, Kinkel RP, Mass MK, Munschauer FE, Priore RL, Pullicino PM, Scherokman BJ, Weinstock-Guttman B, Whitham RH, and The Multiple Sclerosis Collaborative Research Group (MSCRG): Intramuscular Interferon-beta1a for disease progression in relapsing multiple sclerosis. *Ann Neurol* 1996;39(3), 285-294
93. Jacobs LD, Cookfair DL, Rudick RA, Herndon RM, Richert JR, Salazar AM, Fischer JS, Goodkin DE, Granger CV, Simon JH, Emrich LJ, Bartoszak DM, Bourdette DN, Braiman J, Brownscheidle CM, Coats ME, Cohan SL, Dougherty DS, Kinkel RP, Mass MK, Munschauer III FE, Priore RL, Pullicino PM, Scherokman BJ, Weinstock-Guttman B, Witham RH, and The Multiple Sclerosis Collaborative Research Group (MSCRG): A phase III trial of intramuscular recombinant interferon beta as treatment for exacerbating-relapsing multiple sclerosis: design and conduct of study and baseline characteristics of patients. *Multiple Sclerosis* 1995; 1: 118-135
94. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJM, Gavaghan DJ, McQuay HJ: Assessing the Quality of Reports of Randomized Clinical Trials: Is Blinding Necessary? *Control Clin Trials* 1996; 17:1-12
95. Kappos L, Moere EW, Schoetzau A, Schweikert K, Barkhof F, Miller D, Guttmann CRG, Weiner HL, Gasperini C, Filippi M, for the Gadolinium MRI Meta-analysis Group: Predictive value of magnetic resonance imaging for relapse rate and changes in disability or impairment in multiple sclerosis: a meta-analysis. *Lancet* 1999; 353:964-968
96. Karussis DM, Meiner Z, Lehmann D, Gomori JM, Schwarz A, Linde A, Abramsky O: Treatment of secondary progressive multiple sclerosis with the immunomodulator linomide: A double-blind, placebo-controlled pilot study with monthly magnetic resonance imaging evaluation. *Neurology* 1996; 47:341-346
97. Kurtzke JF: Rating neurologic impairment in multiple sclerosis: an Expanded Disability Status Scale (EDSS). *Neurology* 1983; 33:1444-1452
98. Lossef NA, Wang L, Lai HM, Yoo DS, Gawne-Cain ML, McDonald WI, Miller DH, Thompson AJ: Progressive cerebral atrophy in multiple sclerosis: a serial MRI study. *Brain* 1996;119:2009-2019
99. McFarland HF, Barkhof F, Antel J, Miller DH: The role of MRI as a surrogate outcome measure in multiple sclerosis. *Multiple Sclerosis* 2002; 8(1):40-51

100. Meade TW, Wald N, Collins R; Bromhead HJ, Goodman NW; O'Toole LB; Altman DG, Moher D, Rennie D: Letters: CONSORT statement on the reporting standards of clinical trials. *BMJ* 1997; 314:1126-1127
101. Meinert CL: Beyond CONSORT: need for improved reporting standards for clinical trials. *JAMA* 1998; 279:1487-1489
102. Miller DH, Barkhof F, Berry I, Kappos L, Scotti G, Thompson AJ: Magnetic resonance imaging in monitoring the treatment of multiple sclerosis: concerted action guidelines. *J Neurol Neurosurg Psychiatry* 1991;54: 683-688
103. Miller DH, Grossmann RI, Reingold SC, McFarland HF: The role of magnetic resonance techniques in understanding and managing multiple sclerosis. *Brain* 1998; 121:3-24
104. Miller DH; Albert PS, Barkhof F, Francis G, Frank JA, Hodgkinson S, Lublin FD, Paty DW, Reingold SC, Simon J: Guidelines for the use of magnetic resonance techniques in monitoring the treatment of multiple sclerosis. *Ann Neurol* 1996; 39:6-16
105. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF, for the QUORUM group: Improving the quality of reports of meta-analyses of randomised controlled trials: the QUORUM statement. *Lancet* 1999; 354:1896-1900
106. Moher D, Dulberg CS, Wells GA: Statistical Power, Sample Size, and Their Reporting in Randomized Controlled Trials. *JAMA* 1994; 272:122-124
107. Moher D, Jones A, Lepage L; The CONSORT Group: Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. *JAMA* 2001; 285:1992-1995
108. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, Tugwell P, Klassen TP: Does the quality of reports of randomised trials affect estimates of interventions efficacy reported in meta-analyses?. *Lancet* 1998; 352:609-613
109. Moher D, Schulz KF, Altman D, for the CONSORT group: The CONSORT Statement: Revised Recommendations for Improving the Quality of Reports of Parallel-Group Randomized Trials. *JAMA* 2001; 285:1987-1991
110. Moher D, Schulz KF, Altman DG, for the CONSORT Group: The CONSORT Statement: Revised Recommendations for Improving the Quality of Reports of Parallel-Group Randomized Trials. *Ann Intern Med* 2001; 134:657-662
111. Moher D, Schulz KF, Altman DG, for the CONSORT Group: The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001; 357:1191-1194
112. Moreau T, Coles A, Wing M, Isaacs J, Hale G, Waldmann H, Compston A: Transient increase in symptoms associated with cytokine release in patients with multiple sclerosis. *Brain* 1996;119:225-237
113. Mosteller F, Gilbert JP, McPeck B: Reporting Standards and Research Strategies for Controlled Trials – Agenda for the Editor. *Contr Clin Trials* 1980; 1:37-58
114. Münchner Forum "Evidence-Based Health Care": Blockseminar Evidence- Based Medicine – Schwerpunkt Therapie. LMU München, 30.6.2001
115. Nauta JJP, Thompson AJ, Barkhof F, Miller DH: Magnetic resonance imaging in monitoring the treatment of multiple sclerosis patients: statistical power of parallel-groups and crossover designs. *J Neurol Sci* 1994; 122:6-14
116. Nederlands Tijdschrift Voor Geneeskunde: [Eine Standardmethode für den Bericht über randomisierte medizinische Wissenschaftsforschung: die Consolidation of the standards of reporting trials]. *Ned Tijdschr Geneesk* 1998; 142:1089-1091
117. O'Riordan JI, Thompson AJ, Kingsley DPE, MacManus DG, Kendall BE, Rudge P, McDonald WI, Miller DH: The prognostic value of brain MRI in clinically isolated syndromes of the CNS: a 10-year follow-up. *Brain* 1998;121:495-503
118. Otten N: Comparison of drug treatments for multiple sclerosis - systematic review. *Technology Report* 1998, S.18

119. Paolillo A, Coles AJ, Molyneux PD, Gawne-Cain M, MacManus D, Barker GJ, Compston DA, Miller DH: Quantitative MRI in patients with secondary progressive MS treated with monoclonal antibody Campath 1H. *Neurology* 1999; 53:751-757
120. Paty DW, Li DKB, UBC MS/MRI Study Group, IFNB Multiple Sclerosis Study Group: Interferon beta-1b is effective in relapsing-remitting multiple sclerosis II. MRI analysis results of a multicenter, randomized, double-blind, placebo-controlled trial. *Neurology* 1993; 43:662-667
121. Perleth M., Antes G. (Hrsg.): MMW Taschenbuch: Evidenz-basierte Medizin - Wissenschaft im Praxisalltag; EBM-Glossar, Ss. 100-108. Verlag Urban u. Vogel, 2., aktualisierte Auflage 1999
122. Pocock SJ, Hughes MD, Lee RJ: Statistical Problems in the Reporting of Clinical Trials. *N Engl J Med* 1987; 317:426-432
123. Polman CH, Dahlike F, Thompson AJ, Ghazi M, Kappos L, Miltenburger C, Pozilli C: Interferon beta-1b in secondary progressive multiple sclerosis-outline of the clinical trial. *Multiple Sclerosis* 1995; 1:S51-54
124. Porzolt F, Sellenthin C: Der sechste Schritt in der Anwendung der Evidence-Based Medicine. *Z ärztl Fortb qualsich* 2000; 94: 619-620
125. Porzolt F, Thurmayr GR: Skriptum zum Seminar "Evidence-Based Medicine – Schwerpunkt Therapie". AG Evidence-Based Health Care, Humanwissenschaftliches Zentrum der LMU, 30.6.2001
126. Poser CM, Paty DW, Scheinberg L, McDonald WI, Davis FA, Ebers GC, Johnson KP, Sibley WA, Silberberg DH, Tourtellotte WW: New diagnostic criteria für multiple sclerosis: guidelines for research protocols. *Ann Neurol* 1983; 13:227-231
127. Pozzilli C, Bastianello S, Koudriavtseva T, Gasperini C, Bozzao A, Millefiorini E, Galgani S, Buttinelli C, Perciaccante G, Piazza G, Bozzao L, Fieschi C: Magnetic resonance imaging changes with recombinant human interferon- β -1a: a short term study in relapsing-remitting multiple sclerosis. *J Neurol Neurosurg Psychiatry* 1996; 611:251-258
128. PRISMS Study Group: Randomised double-blind placebo-controlled study of interferon beta-1a in relapsing/remitting multiple sclerosis. *Lancet* 1998; 352:1498-1504
129. Rice GPA for the Cladribine Clinical Study Group, Filippi M, Comi G for the Cladribine MRI Study Group: Cladribine and progressive MS. Clinical and MRI outcomes of a multicenter controlled trial. *Neurology* 2000; 54:1145-55
130. Rosati G: Descriptive Epidemiology of Multiple Sclerosis in Europe in the 1980s: A Critical Overview. *Ann Neurol* 1994; 36(S2):S164-S174
131. Rudick R, Antel J, Confavreux C, Cutter G, Ellison G, Fischer J, Lublin F, Miller A, Petkau J, Rao S, Reingold S, Syndulko K, Thompson A, Wallenberg J, Weinshenker B, Willoughby E: Recommendations from the National Multiple Sclerosis Society Clinical Outcomes Assessment Task Force. *Ann Neurol* 1997; 42:379-382
132. Ruiz-Canela M, Martinez-Gonzalez MA, de Irala-Estevez J: Intention-to-treat analysis is related to methodological quality. *BMJ* 2000; 320:1007
133. Sackett DL, Richardson WS, Rosenberg W, Haynes RB: Evidence-based Medicine – How to Practice & Teach EBM. Harcourt Brace Publishers International, London 1996
134. Sackett DL, Richardson WS, Rosenberg W, Haynes RB: Evidenzbasierte Medizin – EBM-Umsetzung und –Vermittlung. Deutsche Ausgabe: Kunz R, Fritsche L. Zuckschwerdt Verlag, Bern, Wien, New York 1999
135. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS: Evidence-Based Medicine: What it is and what it isn't. *BMJ* 1996; 312: 71-72
136. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS, Perleth M (Übersetzer): Was ist Evidenz-basierte Medizin und was nicht? *Münch med Wschr* 1997; 139: 644-645

137. Sailer M, O'Riordan JI, Thompson AJ, Kingsley DP, MacManus DG, McDonald WI, Miller DH: Quantitative MRI in patients with clinically isolated syndromes suggestive of demyelination. *Neurology* 1999; 52:599-606
138. Schäfer H, Berger J, Biebler KE, Feldmann U, Greiser E, Jöckel KH, Michaelis J, Neiß A, Raspe HH, Robra BP, Schumacher M, Trampisch HJ, Victor N, Windeler J: Empfehlungen für die Erstellung von Studienprotokollen (Studienplänen) für klinische Studien. *Informatik, Biometrie und Epidemiologie in Medizin und Biologie* 1999; 30(3):141-154
139. Schulz KF, Chalmers I, Grimes DA, Altman DG: Assessing the Quality of Randomization From Reports of Controlled Trials Published in Obstetrics and Gynecology Journals. *JAMA* 1994; 272:125-128
140. Schulz KF, Chalmers I, Hayes RJ, Altman DG: Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; 273:408-412
141. Schulz KF: Randomised trials, human nature, and reporting guidelines. *Lancet* 1996; 348:596-598
142. Schulz KF: The quest for unbiased research: randomized clinical trials and the CONSORT reporting guidelines. *Ann Neurol* 1997; 41:569-573
143. Schwid SR, Trotter JL: Lessons from Linomide - a failed trial, but not a failure. *Neurology* 2000; 54:1716-1717
144. Simon JH, Jacobs L, Cookfair R, Rudick R et al.: The Natural History of MS Based on an Annual MR Snapshot: Results from the MSCRG Study of Intramuscular Recombinant Interferon beta-1a. *Neurology* 1995; 45(Suppl.4): A418
145. Sipe JC, Knobler RL, Braheny SL, Rice GPA, Panitch HS, Oldstone MBA: A neurologic rating scale (NRS) for use in multiple sclerosis. *Neurology* 1984; 34:1368-1372
146. Smith ME, Stone LA, Albert PS, Frank JA, Martin R, Armstrong M, Maloni H, McFarlin DE, McFarland HF: Clinical worsening in multiple sclerosis is associated with increased frequency and area of gadopentate dimeglumine-enhancing magnetic resonance imaging lesions. *Ann Neurol* 1993;33:480-489
147. Stone LA, Frank JA, Albert PS, Bash C, Smith ME, Maloni H, McFarland HF: The effect of Interferon- β on blood-brain barrier disruptions demonstrated by contrast-enhanced magnetic resonance imaging in relapsing-remitting multiple sclerosis. *Ann Neurol* 1995; 37:611-617
148. Stroup DF, Berlin IA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB, for the Meta-analysis Of Observational Studies in Epidemiology (MOOSE) Group: Meta-analysis of observational studies in epidemiology: a proposal for reporting. *JAMA* 2000; 283:2008-2012
149. The IFNB Multiple Sclerosis Study Group and the University of British Columbia MS/MRI Analysis Group: Interferon beta-1b in the treatment of multiple sclerosis: final outcome of the randomized controlled trial. *Neurology* 1995; 45:1277-1285
150. The IFNB Multiple Sclerosis Study Group: Interferon beta-1b is effective in relapsing-remitting multiple sclerosis. I. Clinical results of a multicenter, randomised, double-blind, placebo-controlled trial. *Neurology* 1993; 43:655-661
151. The Once Weekly Interferon for MS Study Group (OWIMS): Evidence of interferon β -1a dose response in relapsing-remitting MS – The OWIMS Study. *Neurology* 1999; 53:679-686
152. The Quality Standards Subcommittee of the American Academy of Neurology: Practice advisory on selection of patients with multiple sclerosis for treatment with Betaseron. Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology* 1994; 44:1537-1540
153. The Standards of Reporting Trials Group: A proposal for structured reporting of randomised controlled trials. *JAMA* 1994; 272:1926-1931
154. Users' Guides to Evidence-based Medicine *JAMA* 1992; 268(17):2420-2425

155. Weinshenker BG, Bass B, Rice GPA, Noseworthy J, Carriere W, Baskerville J, Ebers GC: The natural history of multiple sclerosis: a geographically based study. I. Clinical course and disability. *Brain* 1989; 112:133-146
156. Weinshenker BG, Rice GP, Noseworthy JH, Carriere W, Baskerville J, Ebers GC: The natural history of multiple sclerosis: a geographically based study. 4. Applications to planning and interpretation of clinical therapeutic trials. *Brain* 1991; 114:1057-1067
157. Williams DH, Davis CE: Reporting Assignment Methods in Clinical Trials. *Contr Clin Trials* 1994; 15:294-298
158. Wötzel C, Wehner C, Pöllmann W: Therapie der Multiplen Sklerose. Ein interdisziplinäres Behandlungskonzept. 2., überarbeitete, Auflage, Pflaum-Verlag München 2000
159. Working Group on Recommendations for Reporting of Clinical Trials in the Biomedical Literature. *Ann Intern Med* 1994; 121:894-895
160. Youl BD, Turano G, Miller DH, Towell AD, MacManus DG, Moore SG, Jones SJ, Barrett G, Kendall BE, Moseley IF, Tofts PS, Halliday AM, McDonald IW: The pathophysiology of acute optic neuritis: an association of gadolinium leakage with clinical and electrophysiological deficits. *Brain* 1991;114:2437-2450

6.3. Sekundärliteratur aus dem Internet:

161. Abteilung Epidemiologie, Sozialmedizin und Gesundheitssystemforschung, Medizinische Hochschule Hannover / Ärztliche Zentralstelle Qualitätssicherung: Discern - Allgemeine Anleitung. IN: <http://www.discern.de/anleitung.htm>, abgefragt am 30.04.02
162. Abteilung Epidemiologie, Sozialmedizin und Gesundheitssystemforschung, Medizinische Hochschule Hannover / Ärztliche Zentralstelle Qualitätssicherung: Discern - Das Discern-Instrument. IN: <http://www.discern.de/instrument.htm>, abgefragt am 30.04.02
163. Centre for Evidence-Based Medicine: Glossary of EBM Terms > More on NNT. IN: <http://www.library.utoronto.ca/medicine/ebm/glossary/nntsPrint.htm>, abgefragt am 05.03.01
164. Centre for Evidence-Based Medicine: Levels of Evidence and Grades of Recommendations. IN: <http://www.minervation.com/ceb2/docs/levels.html>; November 1998, abgefragt am 19.9.02
165. Falck-Ytter Y, Blümler A, Schulgen G, Antes G: Das CONSORT-Statement. IN: <http://www.chocrane.de/deutsch/cccons1.htm>, abgefragt am 23.05.02
166. <http://www.cochrane.org/cochrane/archieco.htm#REF1>
167. http://www.dmsg.de/dmsg/Datenbank/AKTIV_193_5.htm, abgefragt am 28.9.2002
168. <http://www.dr-wolschner.de/html/immun-repair.html>, abgefragt am 28.9.2002
169. <http://www.e-consulters.de/kufri/guete.htm>, abgefragt am 19.9.02
170. <http://www.fhs.mcmaster.ca/ceb/who/faculty/guyatt.htm>
171. <http://www.joannabriggs.edu.au/about.html>: History and Development of the International Evidence Based Practice Movement. Abgefragt am 09.04.2002
172. <http://www.jr2.ox.ac.uk/bandolier/band58/b58-2.html>: NNTs for multiple sclerosis. Abgefragt am 06.06.2002
173. <http://www.jr2.ox.ac.uk/bandolier/band97/b97-3.html>: Mindstretcher 1 - quality and size. Abgefragt am 14.06.02
174. http://www.muskl.de/MS_a/MS_b/frameset.html, abgefragt am 10.01.2001
175. <http://www.ncbi.nlm.nih.gov/PubMed/>, abgefragt zu verschiedenen Zeitpunkten
176. http://www.ruhr-uni-bochum.de/rubin/rbin2_97/rubin5.htm
177. http://www.simmformation.de/html/multiple_sklerose.html, abgefragt am 28.9.2002

178. <http://www.usc.edu/hsc/gcrc/projects/924/newforms/ambulationindex.html>: Protocol 924 Screening - Ambulation Index. Abgefragt am 06.07.2001
179. Kappos L, Comi G, Panitch H, Oger J, Antel J, Conlon P, Steinman L: Multiple Sclerosis: NBI-5788 IN: <http://www.neurocrine.com>, abgefragt am 10.1.2002
180. Köbberling J: Wissenschaft verpflichtet. Eröffnungsvortrag des Vorsitzenden des 103. Kongresses der Deutschen Gesellschaft für Innere Medizin, Wiesbaden, 6. April 1997. IN: <http://neuro24.de/glossare.htm>, abgefragt am 30.05.2002
181. Moher D, Schulz KF, Altman DG: Consensus statement – The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomised trials. BMC Medical Research Methodology 2001; 1:2. IN: <http://www.biomedcentral.com/1471-2288/1/2>
182. Mühlhauser I: UKPDS - Darstellung nach Evidence-based Medicine Kriterien. IN: http://www.uni-duesseldorf.de/WWW/MedFak/MDN/Forum_Kritische_Diabetologie/UKPDSebm.htm, abgefragt am 18.06.02
183. Porzolt F, Richter C, Ehrenstein T, Schouli J, Camara O, Glaubitz R, Goettler S, Hindenburg H-J, Krause U, Merholz A, Schmider A, Stengel D, Walle U, Thim A, Sellenthin C: Essentiell für die Bewertung von Tests in der Medizin: Was ist unter "Likelihood ratio" zu verstehen? IN: http://www.uni-ulm.de/cebm/Publikationen/Likelihood_ratio/likelihood_ratio.html, abgefragt am 25.07.02
184. Porzolt F, Sellenthin C: Der Sechste Schritt in der Anwendung der Evidence-Based Medicine. IN: http://www.uni-ulm.de/cebm/Publikationen/Sechster_Schritt/sechster_schritt.html, abgefragt am 25.07.2002
185. Porzolt F, Thim A, Stengel D, Sehoul J: Wie ist ein gutes Abstrakt zu schreiben und zu präsentieren? IN: <http://www.uni-ulm.de/cebm/Publikationen/Abstrakt/abstrakt.html>, abgefragt am 25.07.2002
186. Raspe H, Stange EF: Evidence Based Medicine: Kontext und Relevanz „Evidenz-gestützter Medizin“. Institut für Sozialmedizin und medizinische Klinik I, Medizinische Universität zu Lübeck. IN: www.qmg.de/qmg/ebm.htm, abgefragt am 15.6.2002
187. Sullivan EM: Healthy Resources for Evidenced-Based Medicine. IN: <http://www.ecmag.net/EC2000/sullivan12.html>, abgefragt am 28.05.2002
188. Weltärztebund: Deklaration von Helsinki, übersetzt von der Bundesärztekammer Außendienst. 48.Generalversammlung des Weltärztebundes in Somerset West 1996. IN: <http://www.ethik.uni-jena.de/Ebene2/Texte/HelsinkiDeklaration96.htm>, abgefragt am 9.11.2000

6.4. In Clegg 2000 und Gold 2000 zitierte Publikationen (falls nicht unter Sekundärliteratur genannt)

189. Barnes MP, Bateman DE, Clelenad PG: Intravenous (IV) methylprednisolone for multiple sclerosis in relapse. J Neurol Neurosurg Psychiatry 1985; 48:157-159
190. Beutler E, Sipe JC, Romine JS, Koziol JA, McMillan R, Zyroff J: The treatment of chronic progressive multiple sclerosis with cladribine. Proc Natl Acad Sci USA 1996; 93:1716-1720
191. Bornstein MB, Miller A, Slagle S, Weitzman M, Crystal H, Drexler E, Keilson M, Merriam A, Wassertheil-Smoller S, Spada V et al.: A pilot trial of Cop 1 in exacerbating-remitting multiple sclerosis. N Engl J Med 1987; 317:408-414.
192. Canadian Cooperative Multiple Sclerosis Study Group. The Canadian cooperative trial of cyclophosphamide and plasma exchange in progressive multiple sclerosis. Lancet 1991; 337:441-446
193. Currier RD, Haerer AF, Meydrech EF: Low dose oral methotrexate treatment of multiple sclerosis: a pilot study. J Neurol Neurosurg Psychiatry 1993; 56:1217-1218.

194. Ebers GC, Hommes O, Hughes RAC, Kappos L, Sandberg-Wollheim M, Palace J: Randomised double-blind placebo-controlled study of interferon beta-1a in relapsing/remitting multiple sclerosis. *Lancet* 1998; 352:1498-1504
195. Goodkin DE, Rudick RA, VanderBrug Medendorp S, Daughtry MM, Schwetz KM, Fischer J, Van Dyke C: Low-dose (7,5mg) oral methotrexate reduces the rate of progression in chronic progressive multiple sclerosis. *Ann Neurol* 1995; 37:30-40
196. Hartung HP, Gonsette R, MIMS- Study Group: Mitoxantrone in progressive multiple sclerosis: A placebo-controlled, randomized, observer-blind phase III trial. *Neurology* 1999; 52(Suppl.2):A290
197. Johnson KP, Brooks BR, Cohen JA, Ford CC, Goldstein J, Lisak RP, Myers LW, Panitch HS, Rose JW, Schiffer RB: Copolymer 1 reduces relapse rate and improves disability in relapsing-remitting multiple sclerosis: results of a Phase III multicenter, double-blind, placebo-controlled trial. *Neurology* 1995; 45:1268-1276
198. Kappos L, Patzold U, Dommasch D, Poser S, Haas J, Krausneck P: Cyclosporine versus Azathioprine in the long-term treatment of MS – results of the German multicenter study. *Ann Neurol* 1988; 23:56-63
199. Kappos L, Polman C, Pozzilli C, Thompson A, Dahlke F: Placebo-controlled multicentre randomised trial of interferon beta-1b in treatment of secondary progressive multiple sclerosis. *Lancet* 1998; 352:1491-1497
200. Killian JM, Bressler RB, Armstrong RM, Huston DP : Controlled pilot trial of monthly intravenous cyclophosphamide in multiple sclerosis. *Arch Neurol* 1988; 45:27-30
201. Milanese C, La Mantia L, Salmaggi A, Eoli M: A double-blind study on azathioprine efficacy in multiple sclerosis: final report. *J Neurol* 1993; 240:295-8
202. Milligan NM, Newcombe R, Compston DAS: A double-blind trial of high dose methylprednisolone in patients with multiple sclerosis: 1. clinical effects. *J Neurol Neurosurg Psych* 1987; 50:511-516
203. Nicholson T, Milne R: Copolymer 1 in relapsing-remitting multiple sclerosis. Southampton: Wessex Institute for Health Research and Development, Development and Evaluation Committee; 1996. Report no.63
204. Romine JS, Sipe JC, Koziol JA, Zyroff J, Beutler E: A double-blind placebo-controlled, randomized trial of cladribine in relapsing-remitting multiple sclerosis. *Proc Assoc Am Physicians* 1999; 111:35-44
205. Sipe JC, Romine JS, Koziol JA, McMillan R, Zyroff J, Beutler E. Cladribine in treatment of chronic progressive multiple sclerosis. *Lancet* 1994; 344:9-13
206. Sørensen PS, Wanscher B, Schreiber K, Blinkenberg M, Jensen CV, Ravnborg M. A double-blind, cross-over trial of intravenous immunoglobulin G in multiple sclerosis: preliminary results. *Multiple Sclerosis* 1997; 3:145-148
207. Steck AJ, Regli F, Ochsner F, Gauthier G: Cyclosporine versus Azathioprine in the treatment of multiple sclerosis: 12- month clinical and immunological evaluation. *Eur Neurol* 1990; 30:224-228
208. The MS Study Group: Efficacy and toxicity of cyclosporine in chronic progressive multiple sclerosis: a randomised, double-blinded, placebo-controlled trial. *Ann Neurol* 1990; 27:591-605
209. Weinstein A, Schwid SIL, Schiffer RB, McDermott MP, Giang DW, Goodman AD: Neuropsychologic status in multiple sclerosis after treatment with glatiramer. *Arch Neurol* 1999; 47:319-324
210. Yudkin PL, Ellison GW, Ghezzi A, Goodkin DE, Hughes RA, McPherson K, Mertin J, Milanese C: Overview of azathioprine treatment in multiple sclerosis. *Lancet* 1991; 338:1051-1055.

7. Verzeichnis der Abbildungen, Tabellen und Diagramme

	Seite
7.1. <u>Abbildungsverzeichnis</u>	
1. Abb. 1: Archie Cochrane, Gründer der Cochrane Collaboration. (www.cochrane.org [166]).	161
2. Abb. 2: Checkliste für MRT-gestützte MS-Therapiestudien inklusive Forderungen aus dem CONSORT-Statement	16
3. Abb. 3: CONSORT-Checkliste zur Publikation randomisierter Studien (http://www.cochrane.de/deutsch/cccons1.htm [166])	19
4. Abb. 4.: Beispiel eines Flussdiagramms für die ein-und ausgeschlossenen Probanden/Patienten im Verlauf einer randomisierten Studie (Aufnahme, Zuordnung, Nachbeobachtung und Datenanalyse). (http://www.cochrane.de/deutsch/cccons1.htm [164])	21
5. Abb. 5: Bewertungsschema nach D. Sackett (Sackett 1999 [134])	22
6. Abb. 6: Beispiel einer Table 1 aus Jacobs, N Engl J Med 2000, S.901 [13]	162
7. Abb. 7: Vergleichstabelle für NNT-Werte verschiedener Erkrankungen. aus: Sackett 1999, S.108 [134]	163
8. Abb. 8: Nomogramm zur Bestimmung von NNT-Werten aus PEER und RRR [134]	164
9. Abb. 9: Objektives Bewertungsschema für die Beurteilung der Validität von Therapiestudien	165
10. Abb. 10: Bewertung einer Studie abhängig von der im objektiven Bewertungsschema erreichten Punktzahl	42
11. Abb. 11: Flussdiagramm aus The Once Weekly Interferon for MS Study Group 1999, S.682 [42]	170
12. Abb. 12: Flussdiagramm aus The Lenercept Multiple Sclerosis Study Group, S.460 [41]	170
13. Abb. 13: Table 1 aus The Lenercept Multiple Sclerosis Study Group, S.460 [41]	171
14. Abb. 14: Tabelle 3 aus Jacobs 1995, S.126 [93]	172
15. Abb. 15: Beschreibung der nicht eingeschlossenen Patienten in Edan 1997, S.114 [4]	173
16. Abb. 16: Tabelle aus The Once Weekly Interferon for MS Study Group 1999, S.684 [42]	173
17. Abbildungen 17 und 18: Levels of Evidence des Oxford Centre for Evidence-based Medicine [164]	174

7.2. Tabellenverzeichnis

1. Tab. 1: Umsetzung der EBM in 5 + 1 Schritten nach Sackett 1997, S.XV [136] und Porzsolt 2001 [125]	9
2. Tab. 2: Untersuchte Studien und deren Themen	13
3. Tab. 3: Bewertete Publikationen und die zur Bewertung herangezogenen Primär-, Sekundär- und Nebenberichte	176
4. Tab. 4: Verwendete Publikation bei Mehrfachberichten der selben Studie	46
5. Tab. 5: Übersicht über die Ergebnisse der Bewertung der eingeschlossenen Studien mithilfe des objektivierten Systems nach Hassler	177
6. Tab. 6: Eckdaten der in das Review eingeschlossenen Studien	179
7. Tab. 7: NNTs der mit mindestens Validitätsstufe 2 bewerteten Studien	182
8. Tab. 8: Auflistung aller Ergebnisse der CONSORT-Auswertung der eingeschlossenen Studien	187
9. Tab. 9, Teil 1: Eckdaten zu den in die Datenerhebung eingeschlossenen Studien (ohne Mehrfachberichte)	189
10. Tab. 9, Teil 2: Auswertungs- und Eckdaten zu den in die Datenerhebung eingeschlossenen Studien (ohne Mehrfachberichte)	191
11. Tab. 10: Verteilung der Anzahl am Ende ausgewerteter Patienten	77
12. Tab. 11: Häufigkeit der vergebenen Punkte in EBM-Schritt 6	85
13. Tab. 12: Gegenüberstellung der klinischen NNT-Ergebnisse für IFN- β aus Berechnungen für die vorliegende Arbeit und aus Clegg 2000 [67] für RRMS-Patienten	91
14. Tab. 13: Gegenüberstellung der klinischen NNT-Ergebnisse für IFN- β aus Berechnungen für die vorliegende Arbeit und aus Clegg 2000 [67] für SPMS-Patienten	92
15. Tab. 14: MRT-NNT-Ergebnisse für IFN- β aus Berechnungen für die vorliegende Arbeit..	93
16. Tab. 15: Evaluationsergebnisse nach dem modifizierten Bewertungsschema der in das Review eingeschlossenen IFN-Studien (Auszug aus Tab. 33)	94
17. Tab. 16: Auflistung der NNT-Ergebnisse für Cladribin aus Berechnungen für die vorliegende Arbeit und Gegenüberstellung mit den Review-Ergebnissen aus Clegg 2000 [67]	99
18. Tab. 17: Auflistung der NNT-Ergebnisse für eine monoklonale Antikörpertherapie aus Berechnungen für die vorliegende Arbeit und Gegenüberstellung mit den Review-Ergebnissen aus Clegg 2000 [67]	101
19. Tab. 18: Klinische und MRT-Ergebnisse für Lenercept aus Berechnungen für die vorliegende Arbeit	103
20. Tab. 19: Klinische und MRT-Ergebnisse für TLI und Steroide aus Berechnungen für die vorliegende Arbeit	104
21. Tab. 20: Klinische und MRT-Ergebnisse für Sulfasalazin aus Berechnungen für die vorliegende Arbeit	106
22. Tab. 21: Klinische und MRT-Ergebnisse für AG284 aus Berechnungen für die vorliegende Arbeit	107
23. Tab. 22: Klinische und MRT-Ergebnisse für Mitoxantron aus Berechnungen für die vorliegende Arbeit	108
24. Tab. 23: Klinische Ergebnisse für Glatirameracetat aus Clegg 2000 [67]	110
25. Tab. 24: Klinische Ergebnisse für Azathioprin aus Clegg 2000 [67]	112
26. Tab. 25: Klinische Ergebnisse für Cyclophosphamid aus Clegg 2000 [67]	114
27. Tab. 26: Klinische Ergebnisse für Methotrexat aus Clegg 2000 [67]	116
28. Tab. 27: Gefundene Häufigkeit der Beschreibung der Randomisationsmethode in früheren Studien im Vergleich mit den Ergebnissen der vorliegenden Arbeit	126
29. Tab. 28: Gefundene Häufigkeit der Beschreibung einer ITT-Analyse in früheren Studien im Vergleich mit den Ergebnissen der vorliegenden Arbeit	130

7.3. Diagrammverzeichnis

1. Diagr. 1: Auswahlsschritte für den Einschluss von Veröffentlichungen in das unsystematische Review.	44
2. Diagr. 2: Erfüllung von CONSORT-Kriterien, Teil 1	50
3. Diagr. 3: Erfüllung von CONSORT-Kriterien, Teil 2	51
4. Diagr. 4: Erwähnung der Randomisation in Titel und Abstract	52
5. Diagr. 5: Eindeutige Kennzeichnung der Zielkriterien	55
6. Diagr. 6: Training der auswertenden Ärzte durchgeführt und beschrieben?	56
7. Diagr. 7: Beschreibung von Zwischenanalysen und Abbruchkriterien	57
8. Diagr. 8: Erfüllung von CONSORT-Forderung Nummer 11: Beschreibung der Verblindung von Patienten und Ärzten	60
9. Diagr. 9: Vorhandensein eines Flussdiagramms	62
10. Diagr. 10: Erfüllung von CONSORT-Forderung Nummer 17: Beschreibung von primären und sekundären Zielkriterien inklusive Konfidenzintervall	67
11. Diagr. 11: Angaben über zusätzliche Tests, insbesondere Subgruppenanalysen	68
12. Diagr. 12: Subgruppenanalysen von Beginn an geplant oder nachträglich durchgeführt?	68
13. Diagr. 13: Verteilung der vergebenen Punkte des objektivierte Bewertungsschemas	193
14. Diagr. 14: Vorläufige Studienbewertung aufgrund der Punktzahl nach den Schritten 1 bis 6 des objektiveren Bewertungsschemas	193
15. Diagr. 15: Häufigkeit der Punktvergabe für die Randomisation alleine	75
16. Diagr. 16: Häufigkeit der Punktvergabe für die zum richtigen Zeitpunkt durchgeführte Randomisation	75
17. Diagr. 17: Häufigkeit der Punktvergabe für die Geheimhaltung der Randomisation	76
18. Diagr. 18: Häufigkeit der Gesamtpunktzahl bei der Bewertung der Randomisation	77
19. Diagr. 19: EBM-Schritt 2: Berücksichtigung aller anfangs eingeschlossenen Patienten.	78
20. Diagramme 20 und 21: EBM-Schritt 3: Häufigkeit der Auswertung nach ITT	79
21. Diagramme 22, 23 und 24: EBM-Schritt 4: Verblindung von Patienten, Therapeuten und Diagnostikern	81
22. Diagr. 25: EBM-Schritt 4: Vergleich der drei Bereiche der Verblindung	81
23. Diagr. 26: EBM-Schritt 5: Häufigkeit der Gleichbehandlung in den Therapiegruppen eingeschlossener Studien	83
24. Diagr. 27: EBM-Schritt 6/1: Vorhandensein einer Tabelle mit den relevanten Patientendaten	28
25. Diagr. 28: EBM-Schritt 6/2: Ähnlichkeit der Vergleichsgruppen	85
26. Diagr. 29: Flussdiagramm für das objektivierte EBM-Bewertungssystem nach Hassler	123
27. Diagr. 30: Flussdiagramm für den Verlauf von Studienteilnehmern während einer Studie. (vgl. Altman 2001, S.666 [48])	129

8. Anhang

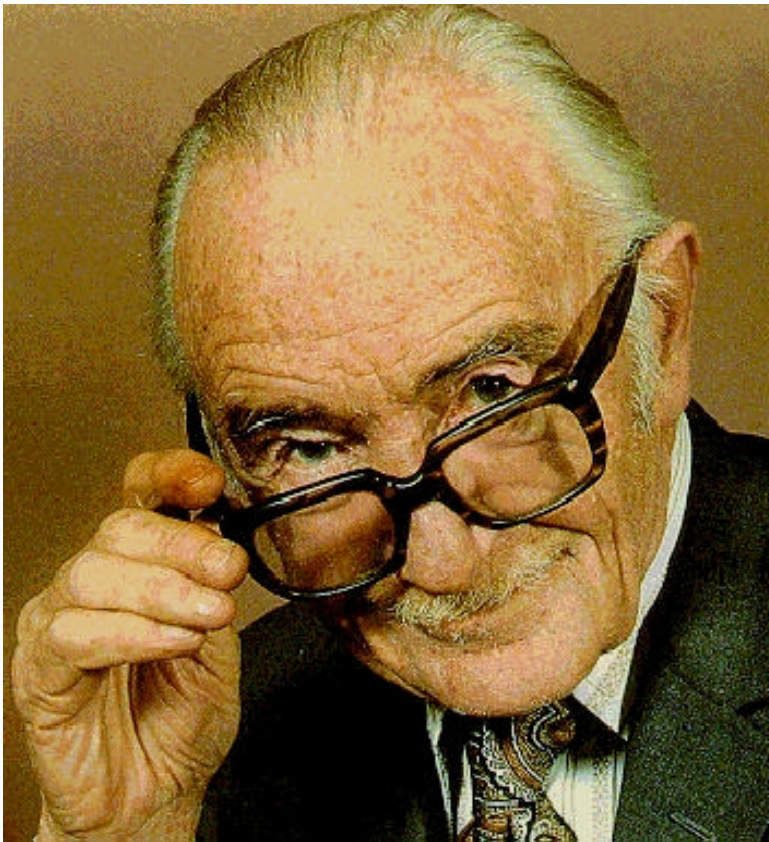


Abbildung 1: Archie Cochrane, Gründer der Cochrane Collaboration.
(www.cochrane.org [166])

TABLE 1. BASE-LINE CHARACTERISTICS OF THE PATIENTS.*

CHARACTERISTIC	ALL PATIENTS (N=383)	INTERFERON BETA-1a GROUP (N=193)	PLACEBO GROUP (N=190)
Female sex – no. (%)	289 (75)	141 (73)	148 (78)
Age – yr	33±7	33±8	33±7
Race or ethnic group – no. (%)			
White	330 (86)	166 (86)	164 (86)
Black	31 (8)	15 (8)	16 (8)
Hispanic	10 (3)	4 (2)	6 (3)
Asian	2 (1)	0	2 (1)
Other	10 (3)	8 (4)	2 (1)
Type of initial event – no. (%)			
Optic neuritis	192 (50)	95 (49)	97 (51)
Spinal cord syndrome	83 (22)	41 (21)	42 (22)
Brain-stem or cerebellar syndrome	108 (28)	57 (30)	51 (27)
Duration of symptoms before initiation of intravenous methylprednisolone – days			
Median	8	8	9
25th and 75th percentiles	5, 12	5, 12	5, 12
Duration of symptoms at initiation of study treatment – days			
Median	19	20	19
25th and 75th percentiles	16, 23	16, 23	16, 23
No. of brain lesions on T ₂ -weighted MRI screening – no. (%)			
2	66 (17)	33 (17)	33 (17)
3-4	125 (33)	64 (33)	61 (32)
5-7	81 (21)	41 (21)	40 (21)
=8	111 (29)	55 (28)	56 (29)
Volume of lesions on base-line T ₂ -weighted MRI – mm ³ †			
Median	2051	2279	1850
25th and 75th percentiles	1003, 4622	1247, 4719	860, 4064
No. of gadolinium-enhancing brain lesions on base-line T ₂ -weighted MRI – no. (%) ††			
0	254 (70)	121 (66)	133 (74)
1	57 (16)	36 (20)	21 (12)
>1	51 (14)	26 (14)	25 (14)
Family history of multiple sclerosis – no. (%)	32 (8)	18 (9)	14 (7)

*Plus-minus values are means ± SD. There were no significant differences between groups. Seven patients (three in the interferon beta-1a group and four in the placebo group) did not meet all eligibility criteria: one patient began intravenous methylprednisolone 15 days after the onset of symptoms, two patients underwent randomization more than 27 days after the onset of symptoms, two patients were older than 50 years at the time of randomization, one patient had pale optic disks at base line, and one patient was found after randomization to have a vascular cause of the initial event rather than demyelination.

† MRI was performed at least four days after the patient completed the course of intravenous methylprednisolone, while the patient was receiving oral prednisone.

†† A total of 362 patients had MRI scans that could be evaluated (183 in the interferon beta-1a group and 179 in the placebo group.)

Abbildung 6: Beispiel einer Table 1 aus Jacobs, N Engl J Med 2000, S.901 [13].

Krankhafter Zustand oder Erkrankung	Intervention	Verhinderte Ereignisse	Ereignisraten		Dauer der Nachbeobachtung	NNT zur Verhinderung eines zusätzlichen Ereignisses
			Kontroll-Zielergebnisrate CER	Experimentelle Zielergebnisrate EER		
Diabetes (IDDM) ^a	Intensivierte Insulinbehandlung	Diabetische Neuropathie	0,096	0,028	6,5 Jahre	15
Diabetes (NIDDM) ^b	Intensivierte + Insulinbehandlung	Verschlechterte diabetische Retinopathie Nephropathie	0,38 0,30	0,13 0,10	6 Jahre	4 5
Akuter Myokardinfarkt ^c	Streptokinase und Aspirin	Tod nach 5 Wochen Tod nach 2 Jahren	0,134 0,216	0,081 0,174	5 Wo. 2 Jahre	19 24
Diastolischer Blutdruck 115–129 mmHg ^d	Antihypertonika	Tod, Schlaganfall oder Myokardinfarkt	0,1286	0,0137	1,5 Jahre	9
Diastolischer Blutdruck 90–109 mmHg ^e	Antihypertonika	Tod, Schlaganfall oder Myokardinfarkt	0,0545	0,0467	5,5 Jahre	128
Selbständige ältere Patienten ^f	Umfassende geriatrische Betreuung zu Hause	Einweisung in ein Dauer-Pflegeheim	0,10	0,04	3 Jahre	17
Schwangere mit Eklampsie ^g	i.v. MgSO ₄ (versus Diazepam)	rezidivierende Krampfstörungen	0,279	0,132	Stunden	7
Gesunde Frauen Alter 50–69 ^h	Brustuntersuchung plus Mammographie	Tod durch Mammakarzinom	0,00345	0,00252	9 Jahre	1075
Hochgradige symptomatische Karotisstenose	Karotisendarteriektomie	Schwerer Schlaganfall oder Tod	0,181	0,08	2 Jahre	10
Frühgeborene ⁱ	Pränatale Kortikosteroidgabe	Atemnotsyndrom	0,23	0,13	Tage	11

Weitere Beispiele finden Sie auf der Homepage dieses Lehrbuchs unter <http://cebim.jr2.ox.ac.uk/> (hier können Sie auch Ihre eigenen Ergänzungen zu unserer Datenbank eintragen).

^a Ann Intern Med 1995; 122: 561-8; EBM 1995; 1: 9. ^d JAMA 1967; 202: 116-22. ^g Lancet 1995; 345: 1455-63; EBM 1996; 1: 44.
^b Diabetes Res Clin Pract 1995; 28: 103-17. ^e BMJ 1985; 291: 97-104. ^h Lancet 1993; 341: 973-8.
^c Lancet 1988; 2: 349-60. ^f N Engl J Med 1995; 333: 1184-9; EBM 1996; 1: 87. ⁱ N Engl J Med 1991; 325: 445-53.
^g Am J Obstet Gynecol 1995; 173: 322-35; EBM 1996; 1: 92.

Abbildung 7: Vergleichstabelle für NNT-Werte verschiedener Erkrankungen. aus: Sackett 1999, S.108 [134].

Beurteilung der Validität von Therapiestudien nach Hassler

Untersuchte Studie:

Zeitschrift	
Jahr, Nummer, Seiten	
Artikel	
Betrifft Arbeitsthese	

1. Wurden die Patienten den Behandlungen randomisiert zugeordnet und wurde die Randomisierungsliste geheimgehalten ?	Randomisierung? zum richtigen Zeitpunkt? Geheimhaltung?	ja/nein ja/nein ja/nein	/3 + /2 + /2 / 7 Punkten Abwertung auf 0 Punkte, wenn überhaupt nicht randomisiert wurde!
2. Wurden bei der Gesamtbeurteilung der Studie alle in die Studie aufgenommenen Patienten berücksichtigt ?	Am Studienende noch vorhandene Patienten (mathematische Rundung): >95% 91-95% 86-90% 81-85% 76-80% 71-75% < 70%		? 6 ? 5 ? 4 ? 3 ? 2 ? 1 ? 0 / 6 Punkten
3. „ Intention-to-treat “-Analyse? ? Wurden die Patienten in den Gruppen analysiert, denen sie zugeordnet waren?	Wird eine Intention-to-treat-Analyse angegeben? <u>ITT-Analyse:</u> -sicher durchgeführt? -eher durchgeführt? -eher nicht durchgeführt? -sicher nicht durchgeführt?	ja/nein	? (1) + ? 4 ? 3 ? 1 ? 0 / 5 Punkten

Februar 2004

Seite 1

Abbildung 9: Objektives Bewertungsschema nach Hassler für die Beurteilung der Validität von Therapiestudien. Einzelheiten siehe Abschnitt 2.5. (fortgesetzt auf der folgenden Seite)

Beurteilung der Validität von Therapiestudien nach Hassler, Seite 2

<p>4. Waren Patienten und Ärzte gegenüber der durchgeführten Behandlung verblindet?</p>	<p>Verblindung der Patienten? Verblindung der behandelnden Ärzte? Verblindung der diagnostizierenden Ärzte?</p>	<p>ja/z.T./nein ja/z.T./nein ja/z.T./nein</p>	<p>/2 + /2 + /2</p>	<div style="border: 1px solid black; width: 100%; height: 100%; position: relative;"> <div style="position: absolute; top: 5px; right: 5px; border-bottom: 1px solid black; width: 80%;"></div> <div style="position: absolute; top: 25px; right: 5px; border-bottom: 1px solid black; width: 80%;"></div> <div style="position: absolute; top: 45px; right: 5px; border-bottom: 1px solid black; width: 80%;"></div> </div> <p style="text-align: center; margin-top: 10px;">/ 6 Punkten</p> <p>Abwertung auf 0 Punkte, falls bei zwei der drei Fragen kein Punkt vergeben werden kann.</p>																																																																																																																																						
<p>5. Wurden die Gruppen, abgesehen von der experimentellen Behandlung, gleich behandelt?</p>	<p><u>Gleichbehandlung:</u> -sicher? -wahrscheinlich? -eher gegeben? -eher nicht gegeben? -sicher nicht?</p>	<p>? 5 ? 4 ? 3 ? 1 ? 0</p>	<div style="border: 1px solid black; width: 100%; height: 100%; position: relative;"> <div style="position: absolute; top: 5px; right: 5px; border-bottom: 1px solid black; width: 80%;"></div> <div style="position: absolute; top: 25px; right: 5px; border-bottom: 1px solid black; width: 80%;"></div> <div style="position: absolute; top: 45px; right: 5px; border-bottom: 1px solid black; width: 80%;"></div> <div style="position: absolute; top: 65px; right: 5px; border-bottom: 1px solid black; width: 80%;"></div> </div> <p style="text-align: center; margin-top: 10px;">/ 5 Punkten</p>																																																																																																																																							
<p>6. Waren die Gruppen zu Beginn der Studie einander ähnlich? Existiert eine Tabelle mit den relevanten Daten?</p>	<p>Gruppenübereinstimmung in allen wichtigen Bereichen? Dazu gehören:</p> <table border="1" style="width:100%; border-collapse: collapse; text-align: center;"> <tr style="background-color: #cccccc;"> <th style="width: 15%;">Die Krankheit ansich</th> <th style="width: 5%;">? ja</th> <th style="width: 5%;">? nein</th> <th colspan="5" style="width: 40%;">0 Punkte gesamt, falls nicht erfüllt</th> <th style="width: 30%;"></th> </tr> <tr style="background-color: #cccccc;"> <th style="text-align: left;">Wichtiges (X/XX) bzw. unwichtiges (0) Kriterium</th> <th>ja</th> <th>eher ja</th> <th>zum Teil</th> <th>eher nein</th> <th>nein</th> <th>k. Ang.</th> <th>Zusatzinformationen zu den jew. Kriterien</th> </tr> <tr style="background-color: #cccccc;"> <td>Anzahl der Patienten pro Behandlungsgruppe</td> <td></td><td></td><td></td><td></td><td></td><td></td><td></td> </tr> <tr style="background-color: #cccccc;"> <td>Schwere der Krankheit</td> <td></td><td></td><td></td><td></td><td></td><td></td><td></td> </tr> <tr style="background-color: #cccccc;"> <td>Alter</td> <td></td><td></td><td></td><td></td><td></td><td></td><td></td> </tr> <tr style="background-color: #cccccc;"> <td>Geschlecht</td> <td></td><td></td><td></td><td></td><td></td><td></td><td></td> </tr> <tr style="background-color: #cccccc;"> <td>Rasse</td> <td></td><td></td><td></td><td></td><td></td><td></td><td></td> </tr> <tr style="background-color: #cccccc;"> <td>Größe</td> <td></td><td></td><td></td><td></td><td></td><td></td><td></td> </tr> <tr style="background-color: #cccccc;"> <td>Gewicht</td> <td></td><td></td><td></td><td></td><td></td><td></td><td></td> </tr> <tr style="background-color: #cccccc;"> <td>Krankheitsentität</td> <td></td><td></td><td></td><td></td><td></td><td></td><td></td> </tr> <tr style="background-color: #cccccc;"> <td>Begleit- und Vorerkrankungen</td> <td></td><td></td><td></td><td></td><td></td><td></td><td></td> </tr> <tr style="background-color: #cccccc;"> <td>Vorbehandlungen</td> <td></td><td></td><td></td><td></td><td></td><td></td><td></td> </tr> <tr style="background-color: #cccccc;"> <td>bisherige Krankheitsdauer</td> <td></td><td></td><td></td><td></td><td></td><td></td><td></td> </tr> <tr style="background-color: #cccccc;"> <td>demographische und geographische Voraussetzungen⁴⁾</td> <td></td><td></td><td></td><td></td><td></td><td></td><td></td> </tr> <tr style="background-color: #cccccc;"> <td rowspan="3">Andere für die vorliegende Studie wichtige Gegebenheiten</td> <td>1)</td><td></td><td></td><td></td><td></td><td></td><td></td> </tr> <tr style="background-color: #cccccc;"> <td>2)</td><td></td><td></td><td></td><td></td><td></td><td></td> </tr> <tr style="background-color: #cccccc;"> <td>3)</td><td></td><td></td><td></td><td></td><td></td><td></td> </tr> </table>			Die Krankheit ansich	? ja	? nein	0 Punkte gesamt, falls nicht erfüllt						Wichtiges (X/XX) bzw. unwichtiges (0) Kriterium	ja	eher ja	zum Teil	eher nein	nein	k. Ang.	Zusatzinformationen zu den jew. Kriterien	Anzahl der Patienten pro Behandlungsgruppe								Schwere der Krankheit								Alter								Geschlecht								Rasse								Größe								Gewicht								Krankheitsentität								Begleit- und Vorerkrankungen								Vorbehandlungen								bisherige Krankheitsdauer								demographische und geographische Voraussetzungen ⁴⁾								Andere für die vorliegende Studie wichtige Gegebenheiten	1)							2)							3)						
Die Krankheit ansich	? ja	? nein	0 Punkte gesamt, falls nicht erfüllt																																																																																																																																							
Wichtiges (X/XX) bzw. unwichtiges (0) Kriterium	ja	eher ja	zum Teil	eher nein	nein	k. Ang.	Zusatzinformationen zu den jew. Kriterien																																																																																																																																			
Anzahl der Patienten pro Behandlungsgruppe																																																																																																																																										
Schwere der Krankheit																																																																																																																																										
Alter																																																																																																																																										
Geschlecht																																																																																																																																										
Rasse																																																																																																																																										
Größe																																																																																																																																										
Gewicht																																																																																																																																										
Krankheitsentität																																																																																																																																										
Begleit- und Vorerkrankungen																																																																																																																																										
Vorbehandlungen																																																																																																																																										
bisherige Krankheitsdauer																																																																																																																																										
demographische und geographische Voraussetzungen ⁴⁾																																																																																																																																										
Andere für die vorliegende Studie wichtige Gegebenheiten	1)																																																																																																																																									
	2)																																																																																																																																									
	3)																																																																																																																																									

Abbildung 9: Fortsetzung von: Objektives Bewertungsschema nach Hassler für die Beurteilung der Validität von Therapiestudien. Einzelheiten siehe Abschnitt 2.5. (fortgesetzt auf der folgenden Seite)

Beurteilung der Validität von Therapiestudien nach Hassler, Seite 3

Vergabe von 0 bis 5 Punkten, je nachdem, ob die für diese Therapie wichtigen Daten übereinstimmen oder nicht.

Empfehlung:

- ? 5 Punkte, wenn fast immer mit „ja“ geantwortet werden kann.
- ? 4 Punkte, wenn überwiegend mit „ja“ geantwortet werden kann.
- ? 3 Punkte, wenn häufig mit „ja“ und ab und zu mit „zum Teil“ geantwortet werden kann.
- ? 2 Punkte, wenn häufiger auch mit „zum Teil“, „nein“ oder „keine Angaben“ geantwortet werden muss.
- ? 1 Punkt, wenn selten mit „ja“ und überwiegend mit „zum Teil“, „nein“ oder „keine Angaben“ geantwortet werden muss.
- ? 0 Punkte, wenn überwiegend mit „nein“ geantwortet werden muss.

? für Vorhandensein einer **Table1** mit Angabe der relevanten Daten: zusätzlich (1) Punkt

/ 6 Punkten

Abschlussbewertung Schritt A:**Übertrag der Einzelpunkte und vorläufige Bewertung**

Nach Beendigung der Bewertung in den Schritten 1-6 erfolgt nun der **Übertrag der Einzelpunkte** aus den Schritten 1) bis 6):

- ? Schritt 1: ___ /7 Punkten
- ? Schritt 2: ___ /6 Punkten
- ? Schritt 3: ___ /5 Punkten
- ? Schritt 4: ___ /6 Punkten
- ? Schritt 5: ___ /5 Punkten
- ? Schritt 6: ___ /6 Punkten



Punktergebnis: ___ / 35 Punkten



Aus dieser Punktzahl folgt die **vorläufige Bewertung:**

	<u>Punkte</u>	<u>Bewertung (Validitätsstufe)</u>
? Voll valide	31 – 35	? 4
? Eingeschränkt valide	28 – 30	? 3
? Eingeschränkt valide	24 – 27	? 2
? Eingeschränkt valide	19 – 23	? 1
? Nicht valide	0 – 18	? 0

Beurteilung der Validität von Therapiestudien nach Hassler, Seite 5

Abschlussbewertung Schritt D:**Synthese****D**

Nach Überprüfung aller maximal möglichen Bewertungen in **A**, **B** und **C** und der daraus ableitbaren Höchstbewertungen folgt dieses endgültige Ergebnis:

Die vorliegende Studie ist:

Voll valide	(4)	?
Eingeschränkt valide	(3)	?
Eingeschränkt valide	(2)	?
Eingeschränkt valide	(1)	?
Nicht valide	(0)	?

Beachte: Unterschiedliche Qualitätsstufen:

Nicht valide:

Falls in einem der sechs Schritte *0 Punkte* erreicht werden, muss die gesamte Studie mit „nicht valide“ (Validitätsstufe 0) bewertet werden. Ein Schritt gilt auch dann als mit 0 Punkten bewertet, wenn die Punkte in Klammern gegeben werden können (Punkte in Klammern zählen sonst jedoch zu der Gesamtpunktzahl dazu). Der Grund dafür ist, dass die entsprechende Studie dann einer anderen Qualitätsstufe angehört und somit nicht mehr den strengen Anforderungen der EBM entspricht.

Eingeschränkte Validität:

Einführung einer weiteren Qualitätsstufe: Wenn in einem der sechs Schritte (ohne Berücksichtigung der Punkte in Klammern in Schritt 3 und 6) nur *2 Punkte oder weniger* gegeben werden können, kann eine Studie nicht mehr als voll valide gelten. Die Maximalwertung ist dann „eingeschränkt valide“ (Validitätsstufe 2).

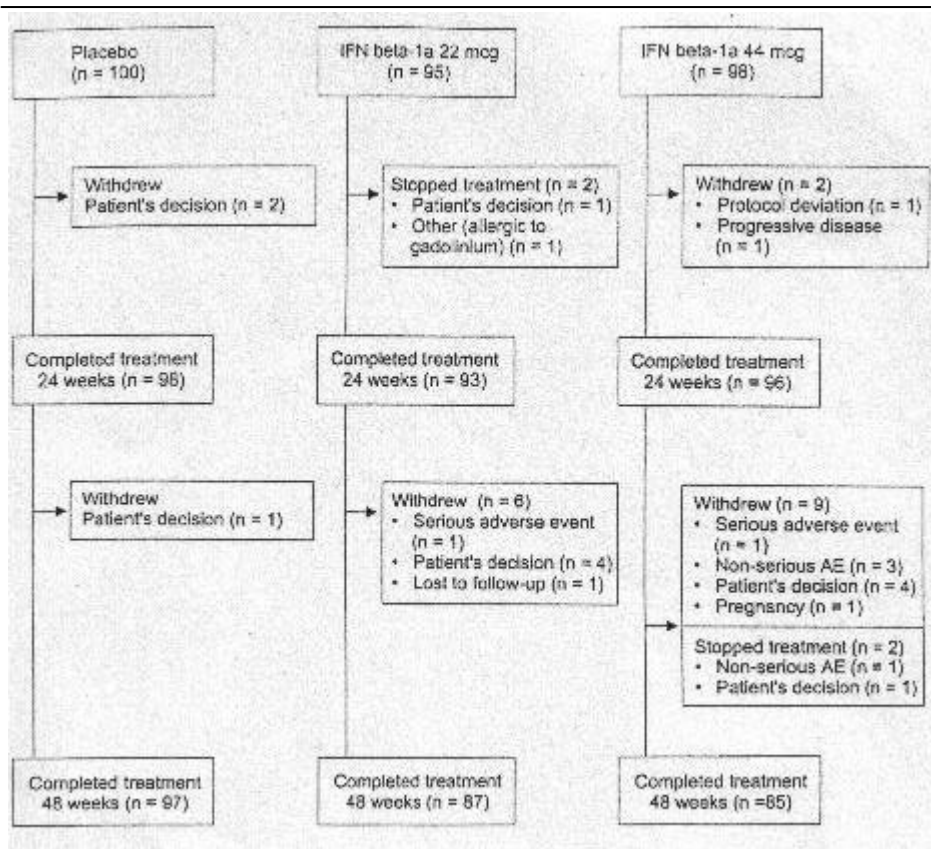


Abbildung 11: Flussdiagramm aus The Once Weekly Interferon for MS Study Group 1999, S.682 [42].

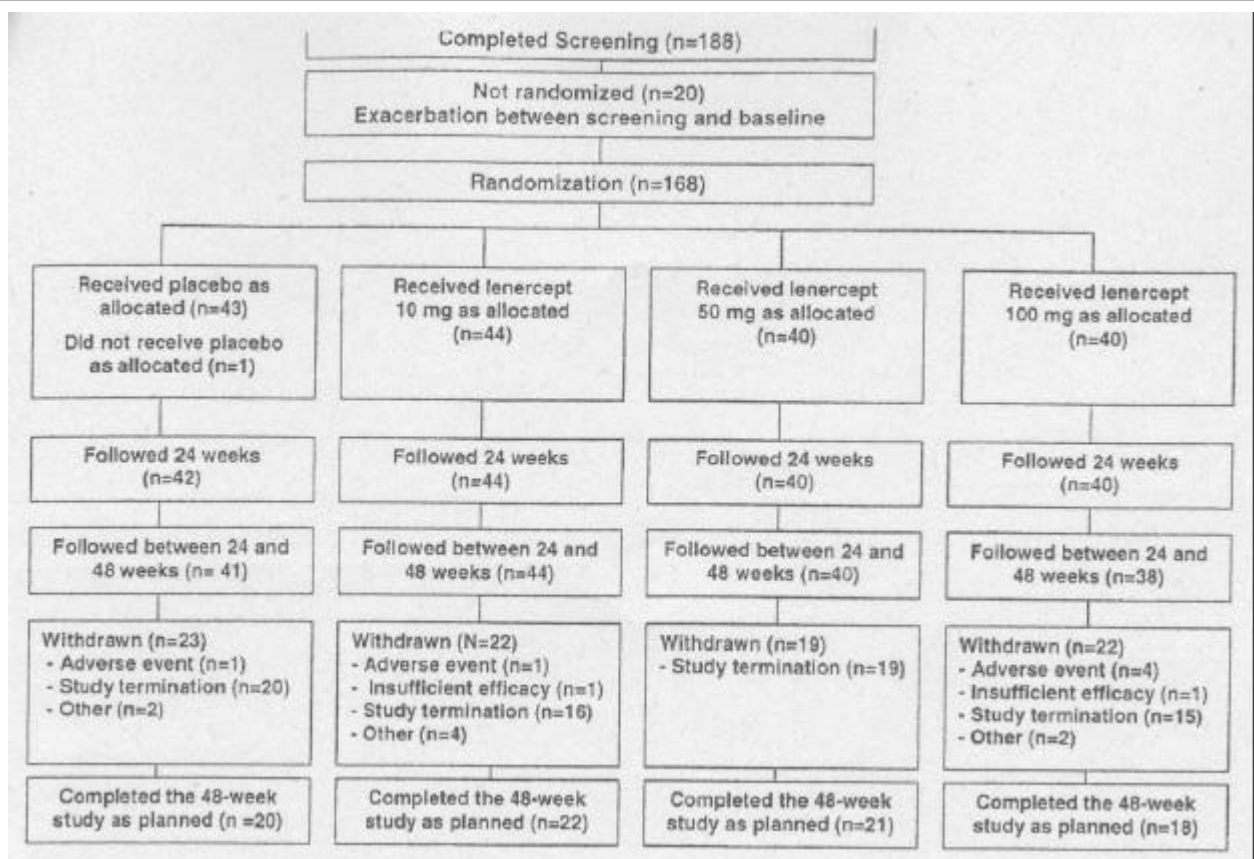


Abbildung 12: Flussdiagramm aus The Lenercept Multiple Sclerosis Study Group, S.460 [41].

Table 1 Demographic and baseline characteristics of patients entered into the lenercept MS trial

Characteristics	Lenercept, mg		
	Placebo, n = 44	10, n = 44	50, n = 40
% Female	66	80	78
Age, y, mean (range)	36.5 (21-50)	34.6 (23-51)	35.1 (19-47)
% White	98	100	93
No. with SPMS	10	5	10
Mean (range) no. exacerbations in prior 2 years	2.7 (2-5)	2.8 (2-8)	2.8 (2-8)
EDSS, mean (range)	2.45 (0-5.5)	2.52 (0-5.0)	2.83 (1.0-5.5)
NRS, mean (range)	83.2 (51-100)	83.7 (44-100)	81.8 (54-100)
			100, n = 40
			73
			34.9 (21-51)
			93
			4
			3.0 (2-6)
			2.55 (0-5.5)
			83.0 (57-99)

SPMS = secondary progressive MS; EDSS = Expanded Disability Status Scale; NRS = Neurological Rating Scale.

Abbildung 13: Table 1 aus The Lenercept Multiple Sclerosis Study Group, S.460 [41].

Table 3 Patient demographics and baseline disease characteristics

<i>Variable</i>	<i>Number</i>	<i>Percent</i>
Patient accrual		
Buffalo	86	29
Cleveland	80	26
Portland	79	26
Washington DC	56	19
Total	301	100
Sex		
Male	80	27
Female	221	73
Race		
Caucasian	278	92
Black	20	7
Other	3	1
Age (years)		
Mean (s.d.)	36.8 (7.4)	
Median	37	
Range	16–55	
Prestudy duration of disease (years)		
Mean (s.d.)	6.5 (5.8)	
Median	4.5	
Range	1–31	
Prestudy exacerbation rate		
Mean (s.d.)	1.2 (0.6)	
Median	1	
Range	0.67–3.7	
Baseline EDSS Score		
Mean (s.d.)	2.3 (0.8)	
Median	2.0	
Range	1.0–3.7	
Baseline MRI characteristics		
Number of patients undergoing MRI		276
Number of gadolinium-enhancing lesions		
Mean (s.d.)	2.76 (6.05)	
Median	1	
Range	0–56	
Percent of patients with Gd-enhancing lesions		53.1
Volume of Gd-enhancing lesions (mm³)^a		
Mean (s.d.)	237.6 (479.2)	
Median	18	
Range	0–2858	
T2 lesion volume (mm³)^a		
Mean (s.d.)	10 431 (10 439)	
Median	7102	
Range	0–61 583	
Percentage of patients with T2 lesions		99.3

^aValues are based on the means of two area determinations

Table 2 Description of the five patients who dropped out*

Time at drop out (months)	EDSS at month -2	EDSS at drop out	No of relapses from month -2	No of new enhancing lesions from month -2	No of scans with new enhancing lesions
4	5.5	6.5	4	35	6/7
4	4.5	5.0	3	14	7/7
4	6.0	8.5	3	98	7/7
3	6.0	7.5	2	51	6/6
5	4.5	8.0	3	99	7/7

*All five patients dropped out because of apparent lack of effectiveness and were in the methylprednisolone group. EDSS = expanded disability status scale.

Abbildung 15: Beschreibung der nicht eingeschlossenen Patienten in Edan 1997, S.114 [4].

Table 5 Percentage of patients ever reporting adverse events during the study

Adverse event	Placebo	Rebif 22 µg	Rebif 44 µg
		qw	qw
Influenza-like symptoms	33	41	60*
Headache	34	48†	50†
Myalgia	11	18	20
Fever	7	8	24*
Chills	3	7	12†
Asthenia	5	2	5
Depression	8	4	8
Injection site inflammation	12	72†	69†
Injection site pain	17	14	17
Skin necrosis	0	0	0
Abnormal liver function test	1	4	4

* Significantly different from placebo and Rebif 22 µg.

† Significantly different from placebo.

Abbildung 16: Tabelle aus The Once Weekly Interferon for MS Study Group 1999, S.684 [42].

Oxford Centre for Evidence-based Medicine Levels of Evidence (May 2001)

Level	Therapy/Prevention, Aetiology/Harm	Prognosis	Diagnosis	Differential diagnosis/symptom prevalence study	Economic and decision analyses
1a	SR (with homogeneity) of RCTs	SR (with homogeneity) of inception cohort studies; CDR validated in different populations	SR (with homogeneity) of Level 1 diagnostic studies; CDR with 1b studies from different clinical centres	SR (with homogeneity) of prospective cohort studies	SR (with homogeneity) of Level 1 economic studies
1b	Individual RCT (with narrow Confidence Interval)	Individual inception cohort study with $\geq 80\%$ follow-up; CDR validated in a single population	Validating cohort study with good reference standards; or CDR tested within one clinical centre	Prospective cohort study with good follow-up	Analysis based on clinically sensible costs or alternatives; systematic review(s) of the evidence; and including multi-way sensitivity analyses
1c	All or none	All or none case-series	Absolute SpPins and SnNouts	All or none case-series	Absolute better-value or worse-value analyses
2a	SR (with homogeneity) of cohort studies	SR (with homogeneity) of either retrospective cohort studies or untreated control groups in RCTs	SR (with homogeneity) of Level >2 diagnostic studies	SR (with homogeneity) of 2b and better studies	SR (with homogeneity) of Level >2 economic studies
2b	Individual cohort study (including low quality RCT; e.g., <80% follow-up)	Retrospective cohort study or follow-up of untreated control patients in an RCT; Derivation of CDR or validated on split-sample only	Exploratory cohort study with good reference standards; CDR after derivation, or validated only on split-sample or databases	Retrospective cohort study, or poor follow-up	Analysis based on clinically sensible costs or alternatives; limited review(s) of the evidence, or single studies; and including multi-way sensitivity analyses
2c	"Outcomes" Research; Ecological studies	"Outcomes" Research		Ecological studies	Audit or outcomes research
3a	SR (with homogeneity) of case-control studies		SR (with homogeneity) of 3b and better studies	SR (with homogeneity) of 3b and better studies	SR (with homogeneity) of 3b and better studies
3b	Individual Case-Control Study		Non-consecutive study; or without consistently applied reference standards	Non-consecutive cohort study, or very limited population	Analysis based on limited alternatives or costs, poor quality estimates of data, but including

					sensitivity analyses incorporating clinically sensible variations.
4	Case-series (and poor quality cohort and case-control studies)	Case-series (and poor quality prognostic cohort studies)	Case-control study, poor or non-independent reference standard	Case-series or superseded reference standards	Analysis with no sensitivity analysis
5	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on economic theory or "first principles"

SR= systematic review. CDR=clinical decision rule. RCT=randomised controlled trial.

Grades of Recommendation

A	consistent level 1 studies
B	consistent level 2 or 3 studies or extrapolations from level 1 studies
C	level 4 studies or extrapolations from level 2 or 3 studies
D	level 5 evidence or troublingly inconsistent or inconclusive studies of any level

"Extrapolations" are where data is used in a situation which has potentially clinically important differences than the original study situation.

Abbildungen 17 und 18: Levels of Evidence des Oxford Centre for Evidence-based Medicine [164].

Bewertete Studien	Nr. laut Literaturverzeichnis	für die Bewertung herangezogene Primär-, Sekundär- oder Nebenberichte
Coles 1999	2	Paolillo, Neurology 1999 [119]
Cook 1997	3	--
Edan 1997	4	Miller 1991 [102]
Filippi, J Neurol Sci 2000	7	Rice 2000* [32]
Filippi, Neurology 2000	8	Rice 2000* [32]
Goodkin 2000	10	--
Jacobs, J Neuroimmunol 2000	12	Jacobs 1996 [92]
Jacobs, NEJM 2000	13	--
Kappos 2000	14	Kappos in: www.neurocrine.com [179]
Kastrukoff 1999	15	The IFNB Multiple Sclerosis Study Group 1993 [150], Paty 1993 [120]
Koch-Henriksen 2000	16	--
Koudriavtseva 1998	17	Pozzilli 1996 [127]
Li 1999	18	PRISMS 1998 [128]
Millefiorini 1997	19	--
Miller 1999	20	European Study Group on Interferon β 1b 1998 [72], Polman 1995 [123]
Molyneux 2000	22	European Study Group on Interferon β 1b 1998 [72]
Myhr 1999	25	--
Nortvedt 1999	26	--
Noseworthy 1998	27	--
Oliveri 1998	28	--
Optic Neuritis Study Group 1997	29	Beck 1992 [61], Beck 1993 [62]
Paolillo, J Neurol 1999	30	--
Patti 1999	31	--
Rice 2000	32	--
Rohowsky-Kochan 1997	33	Cook 1997* [3]
Rudick 1999	34	Jacobs 1995 [93]
Sailer 2000	35	--
Simon 1998	36	Jacobs 1995 [93], Simon 1995 [144], Jacobs 1996 [92]
Sørensen 1998	37	Sørensen 2000* [38]
Sørensen 2000	38	Achiron 1998 [47], Sørensen 1998* [37]
Tan 2000	39	--
Teksam 2000	40	--
The Lenercept Multiple Sclerosis Study Group 1999	41	--
The Once Weekly Interferon for MS Study Group 1999	42	--
Tubridy 1999	43	--
van Oosten 1997	44	--
Wolinsky 2000	45	--
Zhao 2000	46	The IFNB Multiple Sclerosis Study Group 1993 [150], The IFNB Multiple Sclerosis Study Group 1995 [149]

Tabelle 3: Bewertete Publikationen und die zur Bewertung herangezogenen Primär-, Sekundär- und Nebenberichte. *) = selbst eine in der vorliegenden Arbeit bewertete Publikation

Übersicht über die Bewertung der 38 ausgewählten Studien mithilfe des objektivierte EBM-Systems nach Hassler.

Studie Nr.	Randomisierung			Berücksicht. eingeschl. Teilnehmer	ITT		Verblindung		
	Randomis.	Zeitpkt.	Geheimhaltg.		Angeg.	Durchgef.	Patienten	Therapeut	Diagn.
2	3	2	0	5	0	1	1	0	1
3	3	2	2	6	1	3	1	1	2
4	3	2	2	4	0	2	0	0	1
7	3	2	1	5	0	1	1	1	1
8	3	2	1	6	0	2	2	1	2
10	3	2	2	6	1	4	2	2	2
12	3	2	1	0	1	2	2	1	2
13	3	2	1	3	1	3	2	2	2
14	3	2	0	0	0	1	2	0	2
15	3	2	1	6	1	3	2	1	1
16	1	1	0	3	0	1	0	0	0
17	3	0	0	3	0	2	0	0	1
18	3	2	2	6	1	3	2	1	2
19	3	2	2	3	1	3	1	0	1
20	3	2	2	6	1	4	2	1	2
22	2	1	2	0	1	3	2	2	1
25	3	2	0	4	1	1	2	1	2
26	2	2	0	4	1	3	1	0	1
27	3	2	0	5	1	3	2	2	2
28	3	2	0	5	0	0	2	1	2
29	3	2	0	4	0	2	0	1	1
30	3	2	0	0	0	1	0	0	2
31	3	2	1	6	0	3	2	0	2
32	3	2	1	6	0	2	2	1	1
33	3	2	2	6	1	2	1	1	2
34	3	2	2	0	1	1	2	2	2
35	2	2	1	3	0	2	2	0	2
36	3	2	2	2	1	4	2	0	2
37	3	2	0	3	1	3	2	2	2
38	3	2	0	1	1	3	2	1	1
39	3	2	1	0	0	0	1	1	2
40	2	2	0	6	0	1	2	0	2
41	3	2	2	6	1	3	2	1	2
42	3	2	2	6	1	4	2	0	2
43	3	2	0	5	1	3	2	2	2
44	3	2	0	6	0	2	1	0	1
45	3	2	1	0	1	3	2	2	2
46	3	0	1	5	1	3	1	1	2

Tabelle 5 (fortgesetzt auf der folgenden Seite): Übersicht über die Ergebnisse der Bewertung der eingeschlossenen Studien mithilfe des objektivierte Systems nach Hassler. Die Ziffern entsprechen den vergebenen Punkten.

keine Abw. = Diese Studie musste nicht abgewertet werden. Zahlen in Klammern: Aufgrund dieses Schrittes der EBM-Bewertung hätte eine Abwertung auf Stufe 2 erfolgen müssen, doch mangelnde Validität in einem anderen Schritt führte zu einer Abwertung auf Validitätsstufe 0.

Die Studiennummern entsprechen denen des Literaturverzeichnisses.

Studie Nr.	Gleichbeh.	Ähnlichkeit		Gesamtpunkt-zahl	Abwertung		End- ergebnis Validitäts- stufe
		Ähnlich	Tab.		auf Stufe	wegen Schritt	
2	3	1	0	17	2	3,6	0
3	3	2	1	27	keine	keine Abw.	2
4	3	4	1	22	0	(3),4	0
7	3	3	1	22	2	3	1
8	3	4	1	27	2	3	2
10	3	3	1	31	keine	keine Abw.	4
12	3	1	0	18	0	2,(6)	0
13	3	5	1	28	keine	keine Abw.	3
14	3	0	0	13	0	2,6	0
15	3	2	1	26	keine	keine Abw.	2
16	1	0	0	7	0	(1),3,4,(5),6	0
17	3	2	0	14	0	(3),4,(6)	0
18	3	3	1	29	keine	keine Abw.	3
19	3	4	1	24	2	4	2
20	4	2	1	30	keine	keine Abw.	3
22	3	2	1	20	0	2	0
25	3	3	1	23	2	3	1
26	3	3	1	21	2	4	1
27	3	3	1	27	keine	keine Abw.	2
28	3	1	1	20	0	3	0
29	1	5	1	20	2	3,4,5	1
30	3	0	0	11	0	2,(3,4),6	0
31	3	5	1	28	keine	keine Abw.	3
32	3	4	1	26	keine	keine Abw.	3
33	3	2	1	26	keine	keine Abw.	2
34	3	4	1	23	0	2	0
35	3	2	1	20	2	3,4	1
36	3	3	1	25	2	2	2
37	3	2	1	24	keine	keine Abw.	2
38	3	1	1	19	2	2,6	1
39	3	0	0	13	0	2,3,6	0
40	3	0	0	18	0	(3),6	0
41	3	3	1	29	keine	keine Abw.	3
42	3	3	1	29	keine	keine Abw.	3
43	3	4	1	28	keine	keine Abw.	3
44	3	2	0	20	2	3,4	1
45	3	4	1	24	0	2	0
46	3	3	1	24	keine	keine Abw.	2

Tabelle 5 (Fortsetzung von S.165): Übersicht über die Ergebnisse der Bewertung der eingeschlossenen Studien mithilfe des objektivierten Systems nach Hassler. Die Ziffern entsprechen den vergebenen Punkten.

keine Abw.= Diese Studie musste nicht abgewertet werden. Zahlen in Klammern: Aufgrund dieses Schrittes der EBM-Bewertung hätte eine Abwertung auf Stufe 2 erfolgen müssen, doch mangelnde Validität in einem anderen Schritt führte zu einer Abwertung auf Validitätsstufe 0.

Die Studiennummern entsprechen denen des Literaturverzeichnisses.

Autor, Zeitschrift, Jahr	Titel der Studie	Studiendesign	Anz. d. Pat.	MS-Form	Studienform	Ø EDSS-Score (Einschl.-krit.)	Nachbeobachtung	Anzahl der involvierten Ärzte
Simon JH et al., Ann Neurol 1998	Magnetic resonance studies of intramuscular interferon beta-1a for relapsing multiple sclerosis, The Multiple Sclerosis Collaborative Research Group	randomisiert, doppelblind, 2-armig, Placebo	301	RMS	Multicenterstudie, Therapie mit Interferon β -1a (Avonex [®])	2,4 (1,0 – 3,5)	bis 2 Jahre	multicenter
Jacobs LD et al., NEJM 2000	Intramuscular interferon beta-1a therapy initiated during a first demyelinating event in multiple sclerosis, CHAMPS Study Group	randomisiert, doppelblind, 2-armig, Placebo	383	RRMS beginnend	Multicenterstudie, Therapie anfänglich mit Kortikosteroiden, dann Interferon β -1a (Avonex [®])	1,0 (erstes demyelinisierendes Ereign.)	3-4 Jahre (Dauer d. Studie)	? für MRI; phys. Unters.: 1Neurol.; Therapie: 1Neurol.
The Once Weekly Interferon for MS Study Group, Neurology 1999	Evidence of interferon beta-1a dose response in relapsing-remitting MS: the OWIMS Study	randomisiert, doppelblind, 3-armig, Placebo	293	RRMS	Multicenterstudie, Therapie mit Interferon β -1a (Rebif [®])	2,6 (0 – 5,0)	24 Wo. (primäre Dauer); verlängert auf 48 Wochen mit Option der Weiterbeob.	multicenter; 1 Diagn. & 1 Ther./Center
Li DK et al., Ann Neurol 1999	Magnetic resonance imaging results of the PRISMS trial: study of subcutan interferon-beta 1a in relapsing-remitting multiple sclerosis; Prevention	randomisiert, doppelblind, 3-armig, Placebo	560	RRMS	Multicenterstudie, Therapie mit Interferon β -1a (Rebif [®])	2,5 (0,0 – 5,0)	2 Jahre (Dauer d. Studie)	2-3 bei Aktivitätsanalyse, 1-2 bei BOD
Zhao GJ et al., Neurology 2000	Effect of interferon beta-1b in MS: assessment of annual accumulation of PD/T2 activity on MRI, UBC MS/MRI Analysis Group and the MS Study Group	prospect. randomisiert, doppelblind, 3-armig, Placebo	372	RRMS	Multicenterstudie, Therapie mit Interferon β -1b (Betase-ron [®])	2,9 (0,0 – 5,5)	5 Jahre (Dauer d. Studie)	multicenter

Tabelle 6, Teil 1: Eckdaten der in das Review eingeschlossenen Studien (fortgesetzt auf den folgenden Seiten).

Miller DH et al., Ann Neurol 1999	Effect of interferon-beta1b on magnetic resonance imaging outcomes in secondary progressive multiple sclerosis: results of a European multicenter, randomized, double-blind, placebo-controlled trial	randomisiert, doppelblind, 2-armig, Placebo	718	SPMS	Europ. Multicenterstudie, mit Interferon β -1b (Betaseron [®])	5,1 (3,0 – 6,5)	3 Jahre (Dauer d. Studie)	multicenter
Patti F et al., Acta Neurol Scan 1999	Natural interferon-beta treatment of relapsing-remitting and secondary-progressive multiple sclerosis patients, a two-year study	randomisiert, doppelblind, 2-armig, Placebo	98	RRMS, SPMS	nIFN- β 6 MIU vs. Placebo	RRMS: 3,1 (0 – 3,5) SPMS: 5,9 (0 – 7,0)	Dauer 24 Mo., dann 6 Mo. FU für EDSS	1 Neurologe f. Unters., 1 Neurologe f. MRT (KEIN Neuro-rad.)
Goodkin DE et al., Neurology 2000	A phase I trial of solubilized DR2:MBP84-102 (AG284) in multiple sclerosis	randomisiert, doppelblind, Placebo	33	SPMS	Phase 1; Therapie mit AG284	5,5 (3,0 – 7,5)	84 Tage (Dauer d. Studie)	multicenter; 1 Diagn. & 1 Ther./Center
Tubridy N et al., Neurology 1999	The effect of anti-alpha4 integrin antibody on brain lesion activity in MS	randomisiert, doppelblind, 3-armig, Placebo	72	RRMS, SPMS	Anti-alpha4 Integrin Antikörper- Therapie	4,8 (2,0 – 7,0)	24 Wo.(Dauer), davon 20 Wo. follow-up.	1 für Ther.& "normale" Diagn., 1 für EDSS und GNDSS
Noseworthy JH et al., Neurology 1998	The Mayo Clinic-Canadian Cooperative trial of sulfasalazine in active multiple sclerosis	randomisiert, doppelblind, 2-armig, Placebo	199	RRMS (N=151), PPMS (N=48)	Multicenterstudie, Sulfasalazine (Immunmodulator)	2,5 (1,0 – 4,0)	3 Jahre (Dauer d. Studie)	körperl.: 1 Neurolog.; 1 Psych.; MRI: 2 Neurorad.
The Lenercept Multiple Sclerosis Study Group, Neurology 1999	TNF neutralization in MS: results of a randomized, placebo-controlled multicenter study	randomisiert, doppelblind, 4-armig, Placebo	168	RRMS	Multicenterstudie, Therapie mit Lenercept	2,6 (0 – 5,5)	bis zu 48 Wo.	multicenter

Tabelle 6, Teil 2: Eckdaten der in das Review eingeschlossenen Studien (fortgesetzt auf der folgenden Seite).

Sørensen PS et al., Neurology 1998	Intravenous immunoglobulin G reduces MRI activity in relapsing multiple sclerosis	randomisiert, doppelblind, Placebo	26	RRMS, SPMS	intravenous immunoglobulin G (IVIg), Crossoverstudie	4,0 (2,0 – 7,0)	15 Monate (Dauer d.Studie)	2 Neurorad. für MRI (unabh. voneinander), 1 Neurophysiol. für CMCi
Rice GP et al., Neurology 2000	Cladribine and progressive MS: clinical and MRI outcomes of a multicenter controlled trial	randomisiert, doppelblind, 3-armig, Placebo	159	chron. progr. MS	Multicenterstudie, Therapie mit Cladribin	5,6 (3,0 – 6,5)	1 Jahr	multicenter; 1 Neurorad. + 1 Techniker, 1 Untersucher pro Zentrum.
Cook SD et al., J Neurol Sci 1997	Modified total lymphoid irradiation and low dose corticosteroids in progressive multiple sclerosis	randomisiert, doppelblind, 2-armig, Placebo	46	PMS	Prednison in absteigender Dosierung b. jedem Pat., dazu total lymphoid irradiation vs. "Placebo-Bestrahlung".	5,7 (3,5 – 6,5)	Dauer 36 Mo.; k.A. zu FU.	mind. 1 Neurologe für Ther.; mind. 1 Neurologe für Diagn.; 1 Neuroradiologe allein für MRT
Millefiorini E et al., J Neurol 1997	Randomized placebo-controlled trial of mitoxantrone in relapsing-remitting multiple sclerosis: 24-month clinical and MRI outcome	randomisiert, doppelblind, 2-armig, Placebo	51	RRMS	Multicenterstudie, Therapie mit Mitoxantron, (Immunmodulator)	3,6 (2,0 – 5,0)	1 Jahr (Dauer insg. 2 Jahre)	4 Neurologen für EDSS (jeder in 2 Zentren); 2-3 Neuroradiologen für MRT

Tabelle 6, Teil 3: Eckdaten der in das Review eingeschlossenen Studien (Fortsetzung der vorhergehenden Seiten).

Studie; unter- suchte Therapie	Aussage	betroffene Patienten	RRR (CI)	ARR (CI)	NNT (ermittelter Wert) (CI)	Chi- Quadrat
Simon 1998; IFN-β1a	Anzahl der Patienten mit Gd- positiven T1-Läsionen nach 1 Jahr Therapie	PL: 64/113 ET: 60/117	0,09 (-0,15 – 0,29)	5,36 (-7,51 – 18,22)	19 (18,7) (5,5 – 8)	0,66 (n.sign.)
Jacobs, NEJM 2000; IFN-β1b	Anzahl der Patienten mit Entwicklung klinisch definitiver MS nach mindestens 22 Monaten	PL: 73/163 ET: 46/163	0,37 (0,15 – 0,53)	16,56 (6,27 – 26,86)	7 (6,0) (3,7 – 16,0)	9,65 (p<0,01)
	Anzahl der Patienten mit Gehir- Läsionen im T1-gewichteten MRT	PL: 48/114 ET: 15/134	0,73 (0,5 – 0,84)	30,91 (20,39 – 41,43)	4 (3,2) (2,4 – 4,9)	31,06 (p<0,01)
	Anteil der aktiven MR- Scans in der LD*- Gruppe	PL: 49/98 ET: 42/93	0,10 (-0,22 – 0,33)	4,84 (-9,31 – 18,99)	21 (20,7) (5,3 – 8)	0,45 (n.sign.)
	Anteil der aktiven MR- Scans in der HD**- Gruppe	PL: 49/98 ET: 32/96	0,33 (0,06 – 0,53)	16,67 (2,99 – 30,34)	6 (6,0) (3,3 – 33,4)	5,54 (p<0,05)
The Once Weekly Interferon for MS Study Group 1999; IFN-β1a	Anzahl der Patienten mit moderaten oder schweren Rückfällen in der LD*- Gruppe	PL: 40/98 ET: 33/93	0,13 (-0,25 – 0,40)	5,33 (-8,42 – 19,09)	19 (18,8) (5,2 – 8)	0,57 (n.sign.)
	Anzahl der Patienten mit moderaten oder schweren Rückfällen in der HD**- Gruppe	PL: 40/98 ET: 31/96	0,21 (-0,15 – 0,46)	8,52 (-4,97 – 22,02)	12 (11,7) (4,5 – 8)	1,52 (n.sign.)

Tabelle 7 (fortgesetzt auf den folgenden Seiten): NNTs der mit mindestens Validitätsstufe 2 bewerteten Studien. (Legende siehe Seite 186)

Li 1999; IFN-β1a	Anteil der Patienten mit T ₂ - Aktivität über die Studiendauer in der LD*-Gruppe	PL: 169/184 ET: 150/185	0,12 (0,04 – 0,19)	10,77 (3,88 – 17,66)	10 (9,3) (5,7 – 25,8)	9,13 (p<0,01)
	Anteil der Patienten mit T ₂ - Aktivität über die Studiendauer in der HD**-Gruppe	PL: 169/184 ET: 126/182	0,25 (0,16 – 0,32)	22,62 (14,83 – 30,40)	5 (4,4) (3,3 – 6,7)	29,93 (p<0,01)
	Anteil der Patienten mit CU- Läsionen- Aktivität in Kohorte 2 in der LD*-Gruppe	PL: 59/66 ET: 44/64	0,23 (0,07 – 0,36)	20,64 (7,07 – 34,21)	5 (4,8) (2,9 – 14,1)	8,41 (p<0,01)
	Anteil der Patienten mit CU- Läsionen- Aktivität in Kohorte 2 in der HD**-Gruppe	PL: 59/66 ET: 40/68	0,30 (0,14 – 0,43)	26,89 (12,90 – 40,89)	4 (3,7) (2,4 – 7,8)	12,94 (p<0,01)
Zhao 2000; IFN-β1b	Anzahl der Patienten mit annual accumulation of lesion activity (AAA) in der LD*- Gruppe	PL: 62/74 ET: 48/68	0,16 (-0,01 – 0,30)	13,20 (-0,51 – 26,90)	8 (7,53) (3,7 – 8)	3,53 (n.sign.)
	Anzahl der Patienten mit annual accumulation of lesion activity (AAA) in der HD**-Gruppe	PL: 62/74 ET: 49/76	0,23 (0,07 – 0,37)	19,31 (5,66 – 32,96)	6 (5,2) (3,0 – 17,7)	7,27 (p<0,01)
Miller 1999; IFN-β1b	Anzahl der Patienten mit aktiven Scans nach 3 Jahren***	PL: 289/345 ET: 225/350	0,23 (0,16 – 0,30)	19,48 (13,13 – 25,83)	6 (5,1) (3,9 – 7,6)	34,24 (p<0,01)
	Anzahl der Patienten mit Rückfällen über zwei Jahre	PL: 18/29 ET: 12/29	0,33 (-0,12 – 0,60)	20,69 (-4,47 – 45,85)	5 (4,8) (2,2 – 8)	2,49 (n.sign.)
Patti 1999; nIFN-β	Anzahl der RRMS- Patienten mit einer Verschlechterung oder Gleichbleiben des EDSS- Scores	PL: 28/29 ET: 14/29	0,50 (0,27 – 0,66)	48,28 (28,91 – 67,64)	3 (2,1) (1,5 – 3,5)	16,92 (p<0,01)
	Anzahl der SPMS- Patienten mit einer Verschlechterung oder Gleichbleiben des EDSS- Scores	PL: 19/20 ET: 12/20	0,37 (0,08 – 0,56)	35,00 (11,50 – 58,50)	3 (2,9) (1,7 – 8,7)	7,03 (p<0,01)

Tabelle 7 (Fortsetzung): NNTs der mit mindestens Validitätsstufe 2 bewerteten Studien. (Legende siehe Seite 186)

Goodkin 2000; Myelin basic proteins	<ul style="list-style-type: none"> Anzahl der Patienten mit mind. einer Exazerbation über die Studiendauer von 84 Tagen 	PL: 3/8 ET: 6/25	0,36 (-0,99 – 0,79)	0,135 (-23,99 – 50,99)	8 (7,4) (2,0 – 8)	0,56 (n.sign.)
		PL: 10/31 ET: 9/37	0,25 (-0,62 – 0,65)	7,93 (-13,56 – 29,43)	13 (12,6) (3,4 – 8)	0,53 (n.sign.)
		PL: 3/31 ET: 14/37	-2,91 (-11,37 – 0,24)	-28,16 (-46,94 – (-9,38))	-4 (-3,6) (-10,7 – 2)	7,13 (p<0,01)
Noseworthy 1998; Sulfasalazin	<ul style="list-style-type: none"> Anzahl aller Patienten mit mindestens einem Rückfall nach 2 Jahren Anzahl aller Patienten mit mindestens einem Rückfall nach 3 Jahren 	PL: 71/91 ET: 66/102	0,17 (0,01 – 0,31)	13,32 (0,73 – 25,90)	8 (7,5) (3,9 – 136,9)	4,14 (p<0,05)
		PL: 71/86 ET: 75/101	0,10 (0,05 – 0,23)	8,30 (-3,4 – 20,01)	12 (12,1) (5 – 8)	0,82 (n.sign.)
The Lenercept Multiple Sclerosis Study Group 1999; Lenercept	<ul style="list-style-type: none"> Anzahl der Patienten mit mindestens einer Exazerbation nach 24 Wochen Studiendauer Anzahl der Patienten mit mindestens einer Exazerbation nach 48 Wochen Studiendauer Anzahl der Patienten mit schlechtestem NRS-Score nach 48 Wochen 	PL: 15/43 ET: 27/40	-0,94 (2,07 – 0,22)	-32,62 (-52,95 – (-12,28))	-4 (-3,1) (-8,1 – (-1,9))	8,82 (p<0,01)
		PL: 22/43 ET: 32/40	-0,56 (-1,18 – 0,12)	-28,84 (-48,25 – (-9,42))	-4 (-3,5) (-10,6 – (-2,1))	7,58 (p<0,01)
		PL: 29/43 ET: 31/40	-0,15 (-0,50 – 0,12)	-10,06 (-29,13 – 9,01)	-10 (-9,9) (-8 – (-3,4))	1,05 (n.sign.)

Tabelle 7 (Fortsetzung): NNTs der mit mindestens Validitätsstufe 2 bewerteten Studien. (Legende siehe Seite 186)

Sørensen 1998; i.v. Immunglobulin G	<ul style="list-style-type: none"> Anzahl der Patienten mit mindestens einer Exazerbation über die Studiendauer 	PL: 14/21 ET: 6/21	0,57 (0,14 – 0,79)	38,10 (10,17 – 66,02)	3 (2,6) (1,5 – 9,8)	6,11 (p<0,05)
	<ul style="list-style-type: none"> Anzahl der Patienten, die in der LD*- Gruppe im Krankheitsverlauf am Ende fortgeschritten sind 	PL: 14/54 ET: 15/53	-0,09 (-1,03 – 0,41)	-2,38 (-19,22 – 14,47)	-43 (42,1) (-8 – (-5,2))	0,08 (n.sign.)
	<ul style="list-style-type: none"> Anzahl der Patienten, die in der HD**- Gruppe im Krankheitsverlauf am Ende fortgeschritten sind 	PL: 14/54 ET: 15/52	-0,11 (-1,07 – 0,40)	-2,92 (-19,90 – 14,06)	-35 (34,2) (-8 – (-5,0))	0,11 (n.sign.)
Cook 1997; LK-Bestrahlung + Kortikosteroide	<ul style="list-style-type: none"> Anzahl der Patienten, deren EDSS- Score sich 12 Monate nach der Therapie um 1 Punkt verschlechtert hatte 	PL: 7/20 ET: 0/21	1,0	35,00 (14,10 – 55,90)	3 (2,9) (1,8 – 7,1)	8,86 (p<0,01)
	<ul style="list-style-type: none"> Anzahl der Patienten, deren EDSS- Score sich nach 24 Monaten um 1 Punkt verschlechtert hatte 	PL: 14/20 ET: 4/21	0,73 (0,31 – 0,89)	50,95 (24,77 – 77,13)	2 (2,0) (1,3 – 4,0)	10,8 (p<0,01)
	<ul style="list-style-type: none"> Anzahl der Patienten mit gleichgebliebenen oder verschlechterten MRI-Ergebnissen 	PL: 15/20 ET: 9/21	0,43 (0,00 – 0,67)	32,14 (3,71 – 60,57)	4 (3,1) (1,7 – 26,9)	4,36 (p<0,05)

Tabelle 7 (Fortsetzung): NNTs der mit mindestens Validitätsstufe 2 bewerteten Studien. (Legende siehe Seite 186)

Millefiorini 1997; Mitoxantron	<ul style="list-style-type: none"> Anzahl der Patienten, deren EDSS- Score sich nach 12 Monaten Therapie um mindestens 1 Punkt verschlechtert hatte 	PL: 6/24 ET: 2/27	0,70 (0,33 – 0,93)	17,59 (-2,35 – 37,54)	6 (5,7) (2,7 – 8)	2,97 (n.sign.)
	<ul style="list-style-type: none"> Anzahl der Patienten, deren EDSS- Score sich nach 24 Monaten Therapie um mindestens 1 Punkt verschlechtert hatte 	PL: 9/24 ET: 2/27	0,80 (0,17 – 0,95)	30,09 (8,35 – 51,84)	4 (3,3) (1,9 – 12,0)	6,80 (p<0,01)
	<ul style="list-style-type: none"> Anzahl aller Patienten mit mindestens einer Exazerbation im ersten Jahr 	PL: 18/24 ET: 8/27	0,60 (0,26 – 0,79)	45,37 (20,94 – 69,80)	3 (2,2) (1,4 – 4,8)	10,47 (p<0,01)
	<ul style="list-style-type: none"> Anzahl aller Patienten mit mindestens einer Exazerbation im zweiten Jahr 	PL: 14/24 ET: 6/27	0,62 (0,17 – 0,83)	36,11 (10,91 – 61,31)	3 (2,8) (1,6 – 9,2)	6,95 (p<0,01)

Tabelle 7: NNTs der mit mindestens Validitätsstufe 2 bewerteten Studien.

*) LD = Low dose **) HD = high dose ***) Angaben über die zeitliche Auswertung nicht eindeutig;

PL = Placebo; ET= Experimentelle Therapie

	Studie Nr.	2	3	4	7	8	10	12	13	14	15	16	17	18	19	20	22
Bewertungskriterium -	Item-Nr.																
Zuordnung (Randomis.) i. Titel u. Zus.fass.	1	kA	A	Ü+A	kA	kA	A	kA	A	Ü+A	kA	A	A	Ü+A	Ü+A	Ü+A	kA
Hintergrund + Begründung	2	j	j	j	j	j	j	j	j	j	j	t	t	j	j	j	j
Einschlusskriterien	3,1	t	j	j	n	t	j	j	j	j	n	j	j	j	j	j	n
Organisation + Orte	3,2	n	n	n	n	n	n	n	t	n	t	t	n	t	t	n	n
Therapiemethoden: wie + wann	4	j	j	t	j	t	j	j	t	t	j	j	j	j	j	t	t
Ziele/ Hypothesen	5	n	j	t	j	t	j	j	j	t	j	j	n	j	j	j	j
Zielkriterien	6,1	n	np	p+s	t	t	p+s	np	p+s	n	n	p+s	t	n	p+s	p+s	n
Patientendatenerhebung; Ziel-/ Outcome-Kriterien	6,2	j	j	j	j	j	j	t	t	j	j	j	j	j	j	j	j
Meth. z Optimierung d. Ergebnisqualit. (hier: Training der Untersucher)	6,22	n	n	n	j	n	n	n	j	n	n	n	n	j	j	n	t
Fallzahlbestimmung	7,1	n	n	n	j	t	n	j	j	n	n	n	n	j	j	j	n
Zwischenanalysen + Abbruchkriterien	7,2	n	j	k	t	n	k	k	t	j	t	t	n	j	k	j	t
Methode der Zufallszuteilung	8	n	j	n	t	t	j	j	j	n	n	n	n	t	j	n	t
Durchführung der Zuteilung	9,1	t	j	j	t	n	j	j	j	n	j	j	n	t	j	n	n
Geheimhaltung	9,2	n	j	j	z	n	j	z	t	n	z	n	n	z	j	z	z
Wer teilte zu, wer nahm auf	10	n	j	n	n	n	t	n	n	n	n	n	n	n	n	n	n
Patientenverblindung	11,1	n	j	j	t	j	j	j	j	t	j	j	n	t	j	j	t
Therapeutenverblindung	11,2	n	j	n	t	n	j	t	j	n	t	n	n	t	j	n	t
Diagnostikerverblindung	11,3	t	j	j	t	j	j	j	j	j	t	j	t	t	j	j	t
Statistische Methoden	12	n	n	t	j	t	n	n	j	j	t	n	t	j	j	j	j
Inhalte des Flussdiagramms	13	n	n	n	n	n	j	n	n	n	n	n	n	j	n	n	n
Angaben über Rekrutierung	14,1	n	n	n	n	n	t	n	j	n	t	t	n	j	n	n	t
Angaben über FU	14,2	n	n	n	j	n	n	n	n	n	j	n	n	t	t	n	t
Charakteristika der Patienten	15	n	j	j	t	t	j	t	j	n	j	n	j	n	j	n	j
Einschluss aller Patienten	16,1	j	j	j	j	j	j	j	j	j	j	j	j	j	j	j	j
ITT	16,2	n	j	n	n	n	j	j	j	n	t	n	n	j	j	j	t
Angaben über primäre Zielkriterien	17,1	n	j	j	j	t	j	j	j	t	j	n	n	j	j	j	j
Angaben über sekundäre Zielkriterien	17,2	3	j	t	j	t	j	j	j	t	j	n	n	j	j	n	3
Angabe von Konfidenzintervallen	17,3	n	n	n	n	n	n	n	n	n	j	n	n	j	j	n	n
Angabe von ARR, RRR, NNT, OR	17,4	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n
Angabe von p-Werten	17,5	j	j	j	j	n	j	j	j	j	j	n	j	j	j	j	j
Subgruppenanalysen	18,1	j	kS	kS	t	kS	kS	t	kS	j	kS	kS	t	j	j	t	kS
Subgr.analysen vorher oder zusätzlich?	18,2	n	kS	kS	n	kS	kS	n	kS	n	kS	kS	n	n	j	n	kS
UAW	19	j	j	j	t	t	j	j	j	j	t	j	n	t	j	n	t
Interpretation + Biasmöglichkeiten	20	j	t	t	j	t	t	t	t	t	j	n	t	j	t	j	j
Kurze Zusammenfassung der Ergebnisse	20,1	j	j	j	j	j	j	j	j	j	j	n	j	j	j	j	j
Mögliche Mechanismen und ...	20,2	j	n	n	j	n	t	j	n	j	j	n	n	j	n	j	j
Vergleich mit anderen Studien	20,3	j	j	j	j	j	j	t	t	j	j	n	j	j	j	j	j
Limitationen inkl. Methoden, um diese Limitationen zu minimieren	20,4	j	j	j	j	t	t	n	t	n	j	n	t	t	j	j	j
Zusammenfassung der Bedeutung der Ergebnisse/ Auswirkungen	20,5	j	j	j	j	j	t	t	j	j	t	j	t	j	j	j	j
Generalisierbarkeit/ externe Evidence	21	t	n	t	t	n	t	t	t	n	t	t	n	j	t	t	t
Interpret.i. Zus.hang mit akt. Forschung	22	j	j	j	j	j	j	t	t	j	j	n	t	j	j	j	j

Tabelle 8 (fortgesetzt auf der folgenden Seite): Auflistung aller Ergebnisse der CONSORT-Auswertung der 38 eingeschlossenen Studien. kA = keine Angaben; A = Erwähnung nur im Abstract; Ü+A = Erwähnung in Überschrift und Abstract; Ü = Erwähnung nur in der Überschrift; j=ja; n=nein; t=teilweise = Kriterium knapp beschrieben, d.h. teilweise erfüllt oder in Schwesterpublikation enthalten; kS= keine Subgruppe in der untersuchten Studie enthalten und damit Kriterium nicht bewertbar; p=primär, np=nur primär, s=sekundär; z= zentrale Zuteilung, d.h. eine Geheimhaltung ist wahrscheinlich; k=keine.

Studie Nr.	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46
Item-Nr.																						
1	A	A	A	A	kA	kA	A	A	kA	kA	A	A	A	A	kA	A	Ü	A	A	Ü+A	kA	A
2	t	j	j	t	n	j	j	j	t	j	t	t	j	j	j	n	j	t	j	j	j	j
3,1	j	j	j	j	j	t	j	j	t	n	n	j	j	n	j	n	j	j	j	j	j	j
3,2	n	t	n	n	t	n	n	t	n	n	n	t	n	n	n	n	n	n	n	t	n	t
4	t	t	j	t	j	t	j	t	t	j	t	j	j	j	t	t	t	j	j	j	t	j
5	j	j	j	j	j	t	j	j	t	t	t	j	j	n	t	t	j	j	j	t	j	j
6,1	p+s	n	p+s	n	np	n	p+s	p+s	n	np	n	t	p+s	p+s	p+s	n	n	p+s	np	p+s	t	n
6,2	j	t	j	t	j	j	j	j	j	j	t	j	j	j	j	t	j	j	j	j	t	j
6,22	n	n	j	n	n	n	n	t	n	n	n	n	n	n	n	n	n	n	n	n	n	n
7,1	j	t	j	n	t	n	n	j	n	n	n	j	t	n	n	n	n	n	n	j	t	n
7,2	k	t	j	n	t	n	n	j	n	t	k	t	k	n	j	n	t	k	k	n	j	j
8	n	n	t	n	n	n	n	j	t	n	n	n	n	n	n	n	j	j	n	t	n	n
9,1	t	n	j	n	n	n	n	j	t	t	n	n	n	n	j	n	j	j	t	t	t	j
9,2	n	n	n	n	n	n	n	n	t	z	z	z	n	n	z	n	j	j	n	n	z	z
10	n	n	n	n	n	n	n	n	t	n	n	n	n	n	n	n	j	j	n	n	n	n
11,1	t	t	j	j	t	n	j	j	t	j	j	j	j	t	n	j	j	j	j	j	j	t
11,2	n	n	j	n	t	n	n	t	t	j	n	n	n	n	n	n	t	n	j	j	j	t
11,3	t	t	j	j	t	j	j	j	t	j	j	j	j	n	j	j	j	j	j	j	j	j
12	t	n	j	j	t	j	t	j	j	j	j	j	t	n	t	t	j	t	t	j	n	j
13	j	j	j	n	n	n	n	n	n	n	n	n	n	n	j	n	j	j	n	n	n	j
14,1	n	n	j	j	n	n	n	t	n	n	n	n	n	n	n	n	n	j	n	n	n	j
14,2	j	n	j	n	t	n	n	t	t	j	n	t	n	n	n	n	t	j	n	n	n	t
15	j	t	j	j	j	t	j	j	t	j	j	n	j	t	n	t	j	j	j	j	j	n
16,1	j	t	j	j	t	j	j	j	t	t	n	n	j	j	j	j	j	j	j	j	t	j
16,2	j	t	j	n	n	n	n	n	t	t	n	t	j	j	j	n	j	j	j	n	j	t
17,1	j	t	j	j	n	j	j	j	j	t	t	j	j	n	n	t	j	j	j	j	t	j
17,2	t	t	j	t	n	t	t	j	j	t	t	3	j	n	n	n	t	j	j	3	t	j
17,3	j	j	n	n	j	n	n	n	n	n	n	n	n	n	n	n	n	n	j	n	j	n
17,4	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n
17,5	j	n	j	j	j	j	j	j	t	j	j	j	j	j	j	t	j	j	j	j	j	j
18,1	kS	j	j	kS	t	kS	kS	j	kS	j	kS	kS	kS	kS	kS	kS	kS	kS	kS	kS	j	kS
18,2	kS	n	n	kS	n	kS	kS	n	kS	n	kS	kS	kS	kS	kS	kS	kS	kS	kS	kS	n	kS
19	j	j	n	j	n	n	j	j	t	n	n	t	j	j	j	n	j	j	j	j	j	n
20	j	n	t	n	j	j	t	n	t	t	j	j	t	t	t	n	t	t	t	j	j	t
20,1	j	j	j	j	j	j	j	j	j	j	j	j	j	j	j	t	j	j	j	j	j	j
20,2	j	t	j	n	t	j	n	n	j	t	j	j	n	n	n	t	j	n	j	j	j	j
20,3	j	t	t	t	j	j	j	j	j	j	j	j	j	j	n	j	n	j	n	t	t	t
20,4	t	n	n	n	t	t	t	t	n	n	t	j	j	t	t	n	n	t	j	j	t	n
20,5	t	n	j	j	j	j	j	n	t	j	j	j	j	j	j	t	j	j	t	j	j	t
21	t	n	t	n	j	n	n	n	n	n	n	t	t	n	t	t	t	j	t	n	t	n
22	j	t	j	t	j	j	j	j	j	t	t	j	j	j	t	j	j	t	j	j	t	t

Tabelle 8 (Fortsetzung von Seite 187): Auflistung aller Ergebnisse der CONSORT-Auswertung der 38 eingeschlossenen Studien. Legende: siehe Seite 187.

Eckdaten der untersuchten Studien										
Studie	Patientenzahl	Männer	Frauen	durchschn. Alter	klinische Unters.	EDSS	Krh.-dauer (Jahre)	Änderg. d. Scores	MS-Typ*	erhaltene Therapie
Wolinsky 2000	718	247	471	46,4	ja	5,2	15,2	ja	1,2	Linomid 1,0mg, 2,5mg, 7,5mg vs. Plac. pro Tag
Goodkin 2000	33	26	7	46,5	ja	5,2	k.A.	k.A.	2	DR2:MBP84-102 ... 0,6; 2,0; 6,0; 20,0; 60,0; 105,0 mg/kg vs. Plac. 3*an Tagen 0,2,4
Kastrukoff 1999	50	12	38	36,7	ja	2	9,3	k.A.	1	IFN-β-1b 1,6MIU, 8MIU vs. Plac. jeden 2.Tag
Tubridy 1999	72	26	46	40,3	ja	4,8	9,8	nein	1,2	Antegren 3 mg/kg vs. Plac. 2* im Abstand 4 Wo.
The Once Weekly Interferon for MS Study Group 1999	293	76	217	35,3	ja	2,6	6,6	k.A.	1	IFN β-1a 22müg, 44müg vs. Plac. 3*Wo.
Rudick 1999	140	34	106	36,5	ja	2,4	6,5	ja	1	IFN β-1b 6MIU vs. Plac. 1*Wo.
Myhr 1999	97	28	69	33,9	ja	2,8	6,9	k.A.	1	IFN a-2a 4,5MIU vs. Plac. 3*Wo.
Zhao 2000	372	259	113	31,1	ja	2,9	4,4	nein	1	IFN-β-1b 1,6MIU, 8MIU vs. Plac. jeden 2.Tag
Optic Neuritis Study Group 1997	388	89	299	32	ja	0	0	ja	5	Methylprednisolon iv 250mg alle 6h für 3d, dann orales Prednison 1mg/kg/d für 11d vs. orales Prednison 1mg/kg/d für 14d vs. Plac.
Simon 1998	301	221	80	36,8	ja	2,3	6,5	ja	4	IFN β-1a 6MIU vs. Plac. i.m. 1*Wo.
Miller 1999	718	279	439	41	ja	5,1	13,1	ja	2	IFN β-1b 8MIU vs. Plac. jeden 2.Tag
Coles 1999	27	0	0	38,4	ja	5,4	12,5	k.A.	2	zusätzl. Campath-1H (alle), zus. humanized anti-CD4-AB vs. Plac., davon wieder je 1/3 Methylpredn. - vorbeh. vs. TNF-a-Rezeptor-Vorbeh. vs. nichts
Li 1999	560	171	389	34,9	ja	2,5	5,3	ja	1	IFN β-1a 6MIU / 12MIU / Plac. 3*Wo.
Noseworthy 1998	199	69	130	28	ja	2,5	5,5	ja	1	Sulfasalazin 500mg(Steigerung pro Woche:500mg, so lange bis 2g) vs. Plac. 1*/Tag
The Lenercept Multiple Sclerosis Study Group 1999	168	44	124	35,2	ja	2,6	k.A.	nein	1	Lenercept 10mg, 50mg, 100mg vs. Plac. alle 4 Wo.
Oliveri 1998	29	8	21	29,3	ja	k.A.	7,1	ja	1	Methylprednisolon 2g vs. 0,5g iv 1*/Tag; keine Plac.gruppe
Teksam 2000	13	1	12	34,5	nein	k.A.	4,5	k.A.	1	IgG 0,4g/kg / Plac. 1*/Monat an 5 aufeinanderfolgenden Tagen
Jacobs NEJM 2000	383	94	289	33	ja	0	0	ja	1	Prednison über 18 Tage in versch. Dosen, dann: IFN β-1a 30müg vs. Plac. i.m. 1*Wo.
Paolillo 1999	68	0	0	31,8	k.A.	2,2	4,5	k.A.	1	IFN β-1a 3MIU vs. 9MIU s.c. 3*Wo.
Millefiorini 1997	51	16	35	29,9	ja	3,6	5,4	ja	1	Methotrexat 8mg/m ² m vs. Plac. 1*Wo.

Tabelle 9, Teil 1 (fortgesetzt auf der folgenden Seite): Eckdaten zu den in die Datenerhebung eingeschlossenen Studien (ohne Mehrfachberichte).

Molyneux 2000	95	49	46	40	ja	5,1	12	k.A.	2	IFN β -1b 4MIU vs.Plac. s.c. für erste 2 Wo., dann Linomid 2,5mg vs. 5mg vs.Plac. 2*/Tag (Reihe 147), Linomid 2,5mg vs.Plac. für 3 Monate, anschl. 2,5mg od. 5mg abh.von Verträglichkeit vs.Plac. 1*/Tag (Reihe 146)
Tan 2000	413	0	0	k.A.	k.A.	k.A.	k.A.	k.A.	1,2	Cladribine 0,07 mg/kg s.c. 1*/Monat an 5 aufeinanderfolgenden Tagen, 6 Durchgänge Verum und 2 Durchgänge Placebo vs. 2 Durchgänge Verum und 6 Durchgänge Placebo
Filippi J Neurol Sci 2000	159	68	91	44,2	ja	5,6	11,3	ja	5	1,2 Anti-CD4-Antikörper 50mg vs.Plac. i.v. 1*/Monat
van Oosten 1997	71	32	39	36,8	ja	5,5	6,5	ja	1,2	Mitoxantron 20mg + Methylprednisolon 1 g vs. 1,2 Methylprednisolon 1g 1*/Monat
Rice 2000	42	16	26	31,8	ja	4,6	6,3	ja	1,2	1,2 natürliches IFN β 6 MIU vs. Plac. 3*/Wo.
Patti 1999	98	40	58	36,7	ja	4,2	11,6	ja	4	Prednison 30mg (0-4Monate), 25mg (5-8Monate), 20mg (9-12Monate), 15mg (13-16Monate), 10mg (17-36Monate) Tagesdosis. Dazu totale Lymphknotenbestrahlung, Gesamtdosis 1930rad. vs. Placebo-Bestrahlung
Cook 1997	46	19	27	41	ja	5,7	11,1	ja	1	1 NBI5788 5mg vs. 20mg vs. 50mg vs. Plac. 1*/Wo.
Kappos 2000	142	46	96	38,2	ja	k.A.	k.A.	k.A.	1,2	1,2 Amantadin 100mg vs. Plac. 2*/Tag
Sailer 2000	24	10	14	39	ja	4	6,5	k.A.	1	1 IFN β -1b 8MIU s.c. 3*/Wo. vs. IFN β -1a s.c. 1*/Wo
Koch-Henriksen 2000, Teil 2	816	261	555	37,9	ja	2,8	6,9	k.A.	1	IgG 1,0g/kg vs.Plac. i.v. 1*/Mo. 2 aufeinanderfolgende Tage; crossover: 6 Mo. Ther, 1,2 3 Mo. Pause, 6 Mo. Ther.
Sorensen 2000, Teil 1	26	9	17	35	ja	3,5	5	k.A.	1	IgG 0,4g/kg /Plac. i.v. einmalig an 5 aufeinanderfolgenden Tagen; anschl. Booster: 1 IgG 0,4g/kg /Plac. i.v. 1*/Monat
Sorensen 2000, Teil 2	40	32	8	32	ja	2,9	4	ja	1	

*) MS-Typ: 1=RRMS, 2=SPMS, 3=PPMS, 4=sonstige 5=k.A.

Tabelle 9, Teil 1 (fortgesetzt von vorheriger Seite): Eckdaten zu den in die Datenerhebung eingeschlossenen Studien (ohne Mehrfachberichte).

Studie	UAW	Häufigkeit der MR-Scans	KM*)	Dosierung**)	Auswerter MR (Anzahl) ***)	Auswerter zentral	Studien-dauer max. Tage (inkl.FU?)	verwendete Scores
Wolinsky 2000	k.A.	alle 3 Mon.	2	0	1		365	EDSS; SNRS
Goodkin 2000	Hypertonie, Hamwegsinfekt, Kopfschmerzen, Schmerzen an der Injektionsstelle	alle 12,5 Tage	2	2	0		84	EDSS; 9-HPT
Kastrukoff 1999	k.A.	alle 6 Wo.	1	1	2		730	EDSS
Tubridy 1999	Müdigkeit, Schlaflosigkeit	alle 2,6 Wo.	2	2	zentral	1	168	EDSS; GNDS
The Once Weekly Interferon for MS Study Group 1999	Influenza-artige Symptome, Kopfschmerzen, Muskelschmerzen, Fieber, Schüttelfrost, Asthenie, Depression, Entzündung und Schmerzen an der Injektionsstelle, Hautnekrose, irreguläre Leberwerte	alle 3 Wo.	2	0	zentral	k.A.	336	EDSS; SNRS
Rudick 1999	k.A.	alle 12 Monate	2	0	zentral		730	EDSS
Myhr 1999	Grippeähnliche Symptome, Anorexie, Übelkeit, Müdigkeit, Schwächegefühl, Haarausfall, irreguläre Leberwerte, Leuko- und Thrombozytopenie, Anämie	alle 30 Tage	2	0	0		365	EDSS; NRS; SF-36Health Survey; GHQ
Zhao 2000	k.A.	alle 12 Monate	1	1	3		1825	EDSS; SNRS; AAA
Optic Neuritis Study Group 1997	k.A.	alle 2,5 Jahre	1	1	zentral	k.A.	1825	EDSS
Simon 1998	Grippeähnliche Symptome, Muskelschmerzen, Asthenie, Schüttelfrost, Fieber u.a.	alle 8 Monate	2	0	zentral	1	730	EDSS
Miller 1999	z.T. Entz./Nekrose der Injektionsstelle, grippeähnl. Symptome, Muskelhypertonie, RR-Anstieg u.a.	alle 12 Monate	2	2	zentral	k.A.	1095	EDSS
Coles 1999	k.A.	alle 30 Tage mit 1/2 Jahr Pause	2	0	0		690	EDSS
Li 1999	z.T. k.A.	2* pro Jahr	2	2	zentral	3	730	EDSS
Noseworthy 1998	k.A.	alle 6 Monate	2	2	zentral	2	1095	EDSS
The Lenercept Multiple Sclerosis Study Group 1999	Kopfschmerzen, Übelkeit, Bauchschmerzen, „hot flushes“ u.a.	alle 4 Wo.	2	2	zentral	k.A.	336	EDSS; NRS
Oliveri 1998	Schlaflosigkeit, Ängstlichkeit, Dispepsie, Knöchelödeme, metallischer Geschmack	alle 12 Tage	2	2	2		60	EDSS
Teksam 2000	k.A.	alle 3 Monate	0	0	2		365	keine
Jacobs NEJM 2000	influenza-like syndrome, Depressionen	alle 6 Monate	2	0	zentral	k.A.	1460	EDSS
Paolillo 1999	k.A.	1*/Monat für 1 Jahr, dann alle 6 Monate	2	2	3		365	keine

Tabelle 9, Teil 2 (fortgesetzt auf der folgenden Seite): Auswertungs- und Eckdaten zu den in die Datenerhebung eingeschlossenen Studien (ohne Mehrfachberichte).

Millefiorini 1997	k.A.	Übelkeit, Blasenentzündung, Respirationstraktinfektion, Kopfschmerz, Diarrhö, sek. Amenorrhö	alle 12 Monate, Subgruppe zusätzlich alle 2 Mo. für 1/2 Jahr und ein Abschlussscan nach 1 Jahr	0	0	3	730	EDSS	
Molyneux 2000	z.T.	Entz./Nekrose der Injektionsstelle, grippeähn. Symptome, Muskelhypertonie, RR-Anstieg u.a.	alle 6 Monate	1	1	1	1170	keine	
Tan 2000	nein	Pericarditis, Pleuritis, Brustentz., Herzinfarkt	baseline und nach 3 Mo.	2	3	2	270	keine	
Filippi J Neurool Sci 2000	nein	Infektion des ob. Respirationstraktes, Muskelschmerz, Purpura, Reaktionen an der Injektionsstelle, Hypertonie, Rückenschmerzen, Harnwegsinfekt Depression, Arthralgien, Rhinitis, Ataxie, Pharyngitis	alle 12 Monate	2	2	2	730	EDSS; SNRS	
van Oosten 1997	nein	Temperaturanstieg, Neurologische Abweichungen, Herzfrequenzsteigerung, Nausea, Synkope	1*/Monat über 9 Mo., dann Monate 12 und 18	2	2	2	365	EDSS	
Rice 2000	ja	Amenorrhoe, Alopezie, Nausea, Asthenie, Infektion des ob. Respirationstraktes, Harnwegsinfektion, neurologische Ereignisse, Tachykardie/Palpitationen, Hepatitis, Kopfschmerzen, Menorrhagie	1*/Monat	2	2	zentral	2	180	EDSS
Patti 1999	ja	Lymphopenie, Veränderungen im weißen BB, GOT/GPT-Erhöhung, Kopfschmerzen, Müdigkeit, Muskelschmerzen, Arthralgien, Asthenie, trockener Mund, psychopatholog. Veränderungen	alle 6 Monate	0	0	zentral	1	900	EDSS
Cook 1997	nein	Amenorrhoe, gastrointestinale Beschwerden, Haarverlust, Spontan- oder erleichterte traumatische Frakturen, Infektionen, Leukopenie, Gewichtszunahme	inges. 2mal (zu Beginn und Ende der Studie)	2	2	zentral	1	1080	EDSS; TFSS; FSS
Kappos 2000	k.A.	Hypersensitivitätsreaktion, keine weiteren Angaben	bei -1, 0 und 4 Monaten	2	2	0	120	EDSS	
Sailer 2000	ja	k.A.	k.A.	0	0	0	0	EDSS; FSS	
Koch-Henriksen 2000, Teil 2	nein	grippeähn. Symptome, Fieber, Hautsymptomatik	k.A.	2	0	1	0	keine	
Sorensen 2000, Teil 1	k.A.	Thrombosen, Ekzeme	alle 30 Tage mit 3 Monaten Pause	2	2	2	456	EDSS	
Sorensen 2000, Teil 2	k.A.	Müdigkeit, Kopfschmerzen, Hautausschlag, Fieber	alle 12 Monate	0	0	1	730	EDSS	

z. T. = zum Teil; k.A. = keine Angaben; *) Kontrastmittel: 1= keins; 2 =Gd-DTPA; 3= ein anderes; 0= k.A.; **) Kontrastmitteldosierung: 1 = kein KM; 2= 0,1 mmol/kg; 3= andere Dosierung; 0= k.A. ***) pro Zentrum, maximal (d.h. wenn bei Unklarheiten dritter Auswerter hinzugezogen wurde, sind 3 Auswerter eingetragten), Techniker zählen dazu.

Tabelle 9, Teil 2 (fortgesetzt von vorheriger Seite): Auswertungs- und Eckdaten zu den in die Datenerhebung eingeschlossenen Studien (ohne Mehrfachberichte).

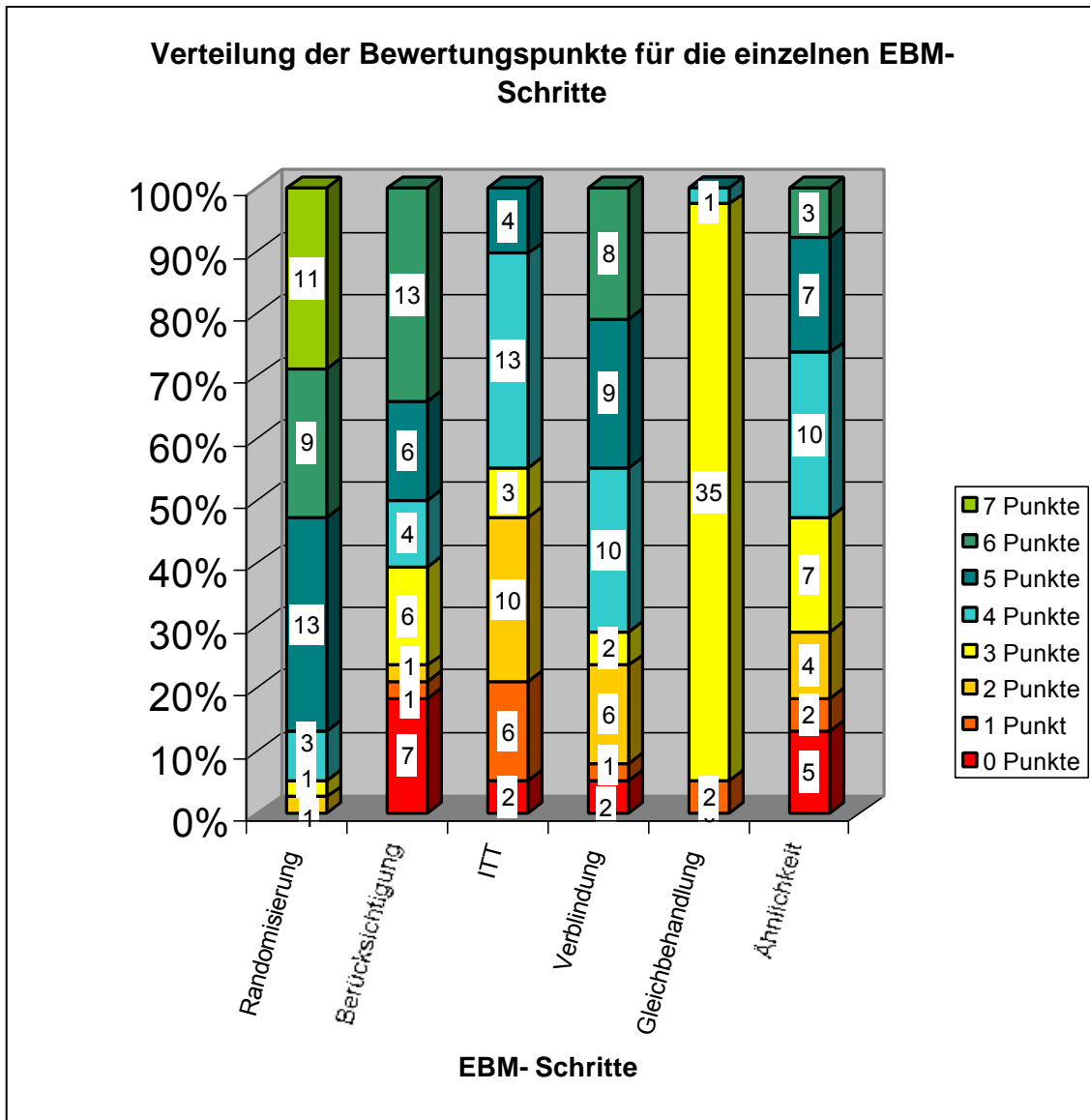


Diagramm 13: Verteilung der vergebenen Punkte des objektivierte Bewertungsschemas. Die Zahlen in den Balken geben die Anzahl der Studien mit der jeweiligen Punktzahl an.

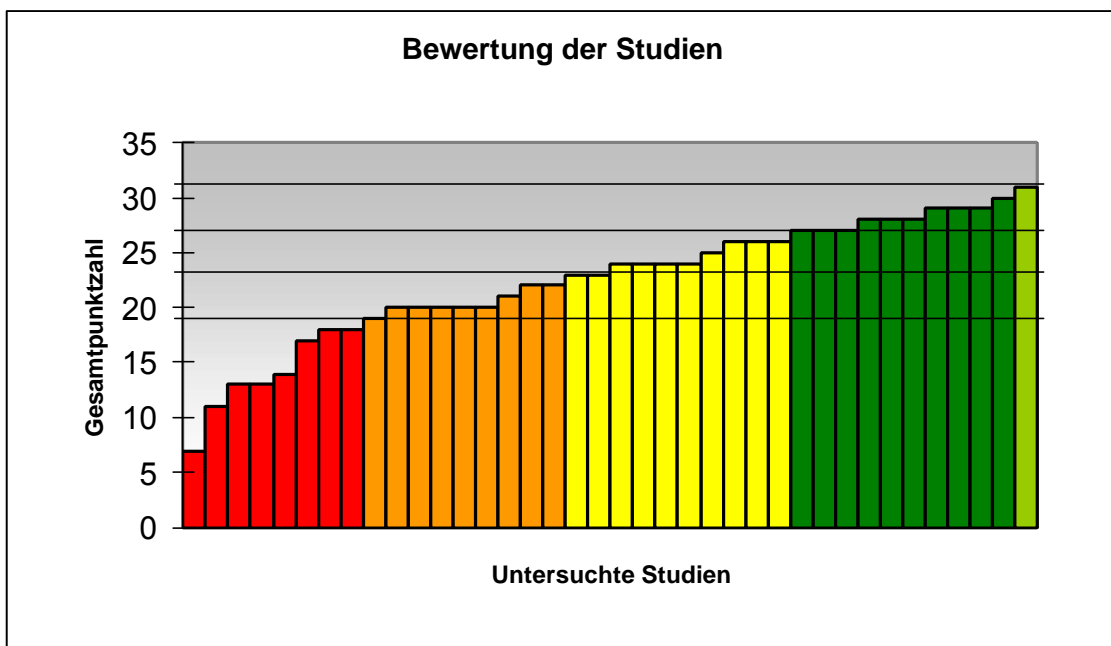


Diagramm 14: Vorläufige Studienbewertung aufgrund der Punktzahl nach den Schritten 1 bis 6 des objektiveren Bewertungsschemas. Die Farbmarkierung gibt die Validitätsgruppen der Studien wieder von rot = Validitätsgruppe 0 bis hellgrün = Validitätsgruppe 4.

Danksagung

Ich danke Herrn Prof. Dr. A. Neiß für die Möglichkeit der Anfertigung meiner Arbeit an seinem Institut.

Des weiteren danke ich Herrn Prof. Dr. F. Porzsolt für die hervorragende Einführung in das Feld der Evidence-based Medicine im Rahmen zahlreicher Kurse und persönlicher Gespräche.

Außerdem danke ich Herrn Prof. Dr. R. Thurmayer für die Unterstützung bei der Erstellung von Graphiken und Tabellen.

Mein ganz besonderer Dank gilt Frau Prof. Dr. G. R. Thurmayer für die intensive persönliche Betreuung und die jederzeit schnelle und kompetente Beantwortung aller Fragen.

Darüber hinaus danke ich meinen Eltern für die immerwährende Unterstützung während der gesamten Zeit meines Studiums.