

Aus dem Institut für Sozialmedizin
der Universität zu Lübeck
Direktor: Prof. Dr. med. Dr. phil. Heiner Raspe

Die Methodik klinischer Forschung im frühen 20. Jahrhundert

Untersuchung der methodischen Gestaltung klinisch-therapeutischer Studien
in Deutschland zwischen 1933 und 1950
anhand einer qualitativen und quantitativen Literaturrecherche

Inauguraldissertation zur Erlangung der Doktorwürde
der Universität zu Lübeck
- Aus der Medizinischen Fakultät der Universität zu Lübeck -

vorgelegt von
Christoph Wolkewitz
aus Eschwege

Lübeck 2009

1. Berichterstatter: Prof. Dr. med. Dr. phil. Heiner Raspe

2. Berichterstatter: Prof. Dr. med. Horst Dilling

Tag der mündlichen Prüfung: 01.07.2010

Zum Druck genehmigt. Lübeck, den 01.07.2010

gez. Prof. Dr. med. Werner Solbach

- Dekan der Medizinischen Fakultät -

Inhaltsverzeichnis

1. Einleitung	S. 5
1.1. Wahrnehmungsprobleme in der Medizin	S. 5
1.2. Die Entwicklung des kontrollierten klinischen Versuchs	S. 7
1.3. Entstehungszusammenhang von Martinis Methodenlehre	S. 11
2. Material und Methoden	S. 16
2.1. Untersuchung der klinischen Forschung in Deutschland	S. 17
2.2. Zur Wahrnehmung von Paul Martini und seiner Methodenlehre	S. 29
2.3. Untersuchung externer Einflüsse	S. 31
3. Untersuchung der klinischen Forschung	
3.1. Darstellung der klinischen Forschung anhand exemplarischer Studien	S. 34
3.1.1. I. Artikel: Das alternierende Verfahren	S. 34
3.1.2. II. Artikel: Nichtinterventionelle Prüfung	S. 38
3.1.3. III. Artikel: Verblindung	S. 39
3.1.4. IV. Artikel: Placebokontrolle	S. 42
3.1.5. V. Artikel: Statistische Auswertung	S. 45
3.1.6. IV. Artikel: Studie aus militärischem Bereich	S. 46
3.2. Quantitative Darstellung der Untersuchungsergebnisse	S. 49
3.2.1. Studienanzahl und Verteilung	S. 49
3.2.2. Studientypen und Kontrollmechanismen	S. 52
3.2.3. Studienplanung und Durchführung	S. 55
3.2.4. Studienauswertung und Darstellung	S. 60
3.3. Zeitliche Entwicklung	S. 66
3.3.1. Untersuchung des Zeitraumes zwischen 1933 – 1950	S. 66
3.3.2. Vergleich der Zeiträume 1918 – 1932 und 1933 – 1950	S. 74

4. Zur Wahrnehmung von Paul Martini und seiner Methodenlehre	S. 78
4.1. Zu den Vorschlägen Paul Martinis in der Methodenlehre	S. 78
4.2. Wahrnehmung von Paul Martinis Ideen und der Methodenlehre	S. 87
4.3. Diskussion um Martinis Methodenlehre	S. 89
5. Untersuchung externer Einflüsse auf die klinische Forschung	S. 98
5.1. Einfluss der Forschung aus dem Ausland	S. 98
5.2. Einfluss aus nicht-zivilen Einrichtung	S. 102
6. Diskussion	S. 111
7. Zusammenfassung	S. 130
8. Anhang	
Literaturverzeichnis	S. 131
Abbildungsverzeichnis	S. 136
Abkürzungsverzeichnis	S. 136
Reichsrichtlinien von 1931	S. 137
Nürnberger Kodex Fassung vom 20.08.1947	S. 140
Auszug aus dem Arzneimittelgesetz	S. 142
Auswertungsübersicht	S. 151
Verschlüsselung der Zeitschriftenrecherche	S. 154
Danksagung	S. 156
Lebenslauf	S. 157

1. Einleitung

1. Wahrnehmungsprobleme in der Medizin

“Nun betrachte, sprach ich, die Lösung und Heilung von ihren Banden und ihrem Unverstande, wie es damit natürlich stehen würde, wenn ihnen folgendes begegnete. Wenn einer entfesselt wäre und gezwungen würde, sogleich aufzustehen, den Hals herumzudrehen, zu gehen und gegen das Licht zu sehen und, indem er das täte, immer Schmerzen hätte und wegen des flimmernden Glanzes nicht recht vermöchte, jene Dinge zu erkennen, wovon er vorher die Schatten sah, was meinst du wohl, würde er sagen, wenn ihm einer versicherte, damals habe er lauter Nichtiges gesehen, jetzt aber dem Seienden näher und zu dem mehr Seienden gewendet, sähe er richtiger, und, ihm jedes Vorübergehende zeigend, ihn frage und zu antworten zwänge, was es sei? Meinst du nicht, er werde ganz verwirrt sein und glauben, was er damals gesehen, sei doch wirklicher als, was ihm jetzt gezeigt werde?“¹

Diese Zeilen entstammen dem Höhlengleichnis aus Platons *Politeia*. In ihr setzt er sich mit dem Erkenntnisproblem auseinander. Als Ursache für das Erkenntnisproblem lässt sich wiederum ein Wahrnehmungsproblem verstehen. Nach Platons Gleichnis unterscheidet sich die Wahrheit gravierend von den Beobachtungen und dem, was für wahrhaftig gehalten wird. Im Zentrum des Gleichnisses steht also die Frage, wie und in welchem Maß der Mensch seiner Wahrnehmung und damit auch seinen Beobachtungen Vertrauen schenken darf. Aus dem Gleichnis lassen sich implizit zwei Voraussetzungen ableiten, die erfüllt sein müssen, um die *wahren Dinge* von den *scheinbar wahren* trennen zu können. Der Betrachter muss sich zunächst den Grenzen der eigenen Wahrnehmung bewusst sein und die bisherigen Beobachtungen kritisch hinterfragen. Hat er seine Beobachtungen einer Prüfung unterzogen und die Grenzen seiner Wahrnehmung akzeptiert, so muss er zweitens die Blickwendung vollziehen, um die *Wahrheit* von der *scheinbaren Wahrheit* unterscheiden zu können.

Dem von Platon beschriebenen Wahrnehmungsproblem begegnet man allerdings nicht nur in Höhlen, sondern – und das macht es an dieser Stelle erwähnenswert - auch am Krankenbett wieder. In der Klinik stellt sich das Wahrnehmungsproblem als nicht minder herausfordernd dar. Das soll kurz verdeutlicht werden.

¹ Eigler 1971: Platon Werke, Band 4, *Politeia* S. 557; Platon erwähnt das Höhlengleichnis in einem politischen Kontext. Das Ideal des hinterfragenden Philosophen, welches Platon beschreibt, lässt sich aber durchaus als allgemeingültig verstehen.

Die Wahrnehmung einer Krankheit ist durch die Anzahl der einzelnen Akteure und durch ihre gegenseitige Beeinflussung ein sehr komplexes Problem im klinischen Alltag. Dabei ist zu unterscheiden zwischen der Wahrnehmung der Erkrankung durch den Patienten auf der einen und durch den Arzt beziehungsweise durch medizinisches Personal auf der anderen Seite. Das Krankheitserleben ist sowohl durch die individuelle psychosoziale Situation als auch durch die physischen Voraussetzungen des Patienten bestimmt. Die Wahrnehmung des Arztes, dessen Aufgabe es ist, den Gesundheitszustand des Patienten möglichst objektiv zu beurteilen, ist wiederum stark von der Empfindung des Patienten beeinflusst. Durch körperliche und technische Untersuchungen versucht der Arzt, sich ein Bild der von dem Patienten geäußerten Empfindungen zu machen. Jedoch sind die Ergebnisse körperlicher Untersuchungen stark von der Qualität und den individuellen Fähigkeiten des Arztes abhängig. So greift er schließlich auf technische Untersuchungen zurück. Doch auch diese stellen nur eine eingeschränkte Möglichkeit dar, klinische Befunde zu objektivieren, da deren Ergebnisse häufig interpretationsabhängig sind und zu unterschiedlichen Bewertungen führen können. So stellt schon die Diagnosefindung ein erhebliches Wahrnehmungs- und Bewertungsproblem im klinischen Alltag dar.

Ist eine Diagnose gefunden und eine Therapie eingeleitet, stellt sich die Frage, wie der Nutzen einer Behandlung zu beurteilen und zu messen ist. Dabei ist eine objektive Beurteilung einer therapeutischen Maßnahme keineswegs als leichter zu betrachten als die Diagnosefindung zuvor. Die Wahrnehmung eines therapeutischen Einflusses findet auf den gleichen Ebenen und unter ähnlichen Voraussetzungen statt, wie zuvor schon beschrieben. Hinzu kommt noch, dass der Arzt, der eine Therapie einleitet, dieses mit dem Ziel verfolgt, eine Heilung oder zumindest eine Linderung der Beschwerden zu bewirken, während der Patient in der Regel auf Heilung hofft. Der Arzt würde also eine Therapie nicht einleiten, würde er nicht von einem Nutzen von dieser ausgehen und der Patient wiederum würde keine Therapie akzeptieren, wenn diese nicht erfolgversprechend ist. Durch diese voreingenommene Haltung sowohl des Patienten als auch des Arztes ist eine objektive Beurteilung der Behandlung durch diese Akteure allein sehr schwierig. Auch die Zuhilfenahme von technischen Parametern, die im Verlauf beobachtet werden und ein scheinbar objektives Urteil erlauben, ist häufig genug nicht befriedigend. Das wiederum liegt an der oft zu beobachtenden Diskrepanz zwischen klinischen und technischen Befunden.

Die Wahrnehmungsprobleme in der Klinik haben also eine Unsicherheit in der Einschätzung therapeutischer Effekte als Folge. Diese Unsicherheit hat ihrerseits dazu

geführt, dass eine große Anzahl von therapeutischen Maßnahmen existiert, deren Einsatz mit individuellen positiven Erfahrungen gerechtfertigt wird, für deren Nutzen es aber keinen zufrieden stellenden Nachweis gibt.

Welche Bedingungen müssen also erfüllt sein, damit das Wahrnehmungsproblem gelöst wird und diese Unsicherheit überwunden werden kann?

Übersetzt man vor diesem Hintergrund die beiden zentralen Aussagen aus dem Höhlengleichnis in diesen Zusammenhang, so könnten diese wie folgt lauten:

1. Wahrnehmungen über den Nutzen einer Therapie sind mit größtmöglicher Skepsis zu begegnen! In einem komplexen System, wie in der Klinik, sind zu viele Einflussfaktoren involviert, die einen unkritischen Umgang bei der Bewertung einer Therapie verbieten!
2. Die von Platon erwähnte Blickwendung bedeutet, dass zur Überwindung des Wahrnehmungsproblems und für einen Erkenntnisgewinn die kompromisslose Anwendung wissenschaftlicher Methoden bei der Beobachtung unabdingbar ist! Der auf das individuelle Patientenschicksal fixierte Blick des Klinikers muss gelockert werden und es muss ein adäquater Beobachtungsrahmen gewählt werden, der eine zuverlässige Beurteilung über Heilungserfolge zulässt!

2. Die Entwicklung des kontrollierten klinischen Versuchs

In der modernen klinischen Forschung werden als Antwort auf dieses Wahrnehmungsproblem verschiedene Ansätze verfolgt, um die Effektivität einer Therapie objektiv messbar zu machen.

Ein Ansatz beinhaltet eine zunehmende Technisierung medizinischer Beobachtungen. Dadurch wird versucht, klinische Verläufe zum Beispiel anhand von Laborparametern zu beschreiben und somit ein objektives Maß über den Therapieerfolg zu generieren. In diesen Technisierungsprozess ist mittlerweile eine Vielzahl von diagnostischen Maßnahmen eingebunden, die von bildgebenden Verfahren wie der Magnetresonanztomographie bis hin zu labortechnischen wie der Polymerase-Chain-Reaction reichen. Der wissenschaftliche Wert dieses Ansatzes ist unbestreitbar, da dieser häufig eine objektive und messbare Darstellung eines klinischen Befundes und somit auch die Beurteilung des Effektes einer Therapie ermöglicht. Dieser Weg allein stellt jedoch nur einen Teil zur Lösung des Wahrnehmungsproblems in der klinischen Forschung dar. Denn trotz aller technischen Innovation liefert dieser Ansatz Ergebnisse, die keineswegs als eindeutig angesehen werden dürfen und Spielraum für Interpretationen lassen. Zudem

suggeriert eine zunehmende Technisierung häufig eine trügerische Sicherheit in Diagnostik und Therapie, die aufgrund des häufig nur lückenhaften Verständnisses biologischer Vorgänge unangemessen scheint.

Ein anderer Weg hat zu einer systematischen und normierten Erfassung klinischer Daten geführt. Diese findet auf verschiedenen Ebenen klinischer Prüfungen statt. Primär werden in klinischen Studien Daten über die Effektivität verschiedener Heilverfahren gewonnen. Sekundär werden diese Primärdaten zum Beispiel in Metaanalysen mit Ergebnissen aus anderen Versuchsreihen verglichen und ausgewertet. Dabei spielen nicht nur die bloßen Fakten wie zum Beispiel Überlebensraten eine Rolle, sondern auch Fragen hinsichtlich der Lebensqualität unter einer gewählten Therapie.

In der primären Datengewinnung werden klinische Studien in verschiedenen Formen durchgeführt. Man kann grob unterscheiden zwischen *kontrollierten* und *nicht-kontrollierten* Studien. Dabei hat sich die Versuchsanordnung in Form der *randomisierten kontrollierten Studie* – kurz RCT - als ein sehr wichtiges methodisches Verfahren in der klinischen Beobachtung in den letzten Jahrzehnten bewährt. Diese Anordnung wurde in der heutzutage üblichen Art erstmals 1948 in dem „Streptomycin Trial“ umgesetzt.² Das für die RCT charakteristische Merkmal stellt die *Randomisierung* bei der Gruppenbildung eines kontrollierten Versuchs dar. Die Probanden einer Studie werden dabei nach dem Zufallsprinzip verschiedenen Gruppen zugewiesen. Diese Zuordnung dient im Wesentlichen dazu, Verzerrungen durch willkürliche Zuteilung nach bestimmten Merkmalen der Probanden - wie zum Beispiel Alter oder Geschlecht – vorzubeugen. Idealerweise gewährleistet diese Art der Gruppenzuweisung, dass die Gruppen in allen für die Studie relevanten Merkmalen vergleichbar sind.³

Oftmals werden, ohne dass dies einen obligaten Bestandteil eines RCT darstellt, die Versuchsprobanden und Untersucher *verblindet*, das heißt, dass weder der Untersucher noch der Teilnehmer weiß, ob sie der Kontroll- oder der Therapiegruppe angehören. Zu diesem Zweck werden häufig auch *Placebos* also Scheinmedikamente eingesetzt.⁴ Diese aufwändige Studiengestaltung ist in einem solchen komplexen wissenschaftlichen Umfeld wie der klinischen Forschung sinnvoll und dient wie die Randomisierung dem Ziel, Verzerrungen zu vermeiden und eine möglichst zuverlässige Beobachtung klinischer Sachverhalte zu gewähren. Auch die Merkmale der *Versuchsverblindung* und der Einsatz

² Vgl Kaptchuk 1998 429: Kaptchuk sieht in dem von Austin Bradford durchgeführten Streptomycin Trial von 1948 einen der ersten Randomisierten Kontrollierten Versuche (S. 429).

³ Fletcher et al 1999 S. 193 ff.

⁴ Fletcher et al 1999 S. 203 ff.

von *Placebokontrollen* – obwohl speziell die Placebokontrolle nicht als unumstritten zu betrachten ist⁵ - lassen sich als prägende Charakteristika moderner klinischer Studien auffassen.

Die RCTs wurden – wie bereits erwähnt – erstmals Ende der 40er Jahre des 20. Jahrhunderts durchgeführt. Erste Ansätze zu kontrollierten Studien sind allerdings nachweislich spätestens 200 Jahre zuvor bereits durchgeführt worden. Eine kurze Darstellung exemplarischer Studien soll einzelne Entwicklungsschritte des klinischen Versuchs im zeitlichen Zusammenhang illustrieren.

1747 wurde durch den Schotten James Lind ein Versuch zur Behandlung der Skorbut durchgeführt.⁶ Wie Lind ausführt, gab es eine Reihe von Überlieferungen und Berichten zur Therapie der Skorbut. Doch bei vielen, so bemängelte er, existierte kein genügender Wirksamkeitsnachweis. So entschloss sich der Schiffsarzt der Royal Navy dazu, eigene Beobachtungen zur Behandlung der Skorbut durchzuführen. Zu diesem Zweck wurden an Bord der MS Salisbury 12 an Skorbut erkrankte Seemänner sechs Gruppen zu je zwei Personen zugewiesen. Die Gruppen wurden sechs verschiedenen Behandlungen unterzogen, lebten aber ansonsten unter gleichen Bedingungen, wie er betonte. Lind kommt nach Durchführung seiner vergleichenden Beobachtungen zu dem Ergebnis, dass Zitronen und Limonen für die Behandlung zu empfehlen seien.

In dem Versuch von Lind begegnet man wichtigen Elementen einer systematischen klinisch-therapeutischen Beobachtung. Zunächst sind die Bedingungen so gewählt, dass der Wirksamkeitsnachweis möglichst frei ist von externen Störfaktoren, wie zum Beispiel Unterschieden in der Ernährung oder des Geschlechts. Dies erreichte Lind, indem er seine Beobachtung an einem relativ homogenen Kollektiv unter ansonsten gleichen Bedingungen durchgeführt hat. Zum anderen führte er eine kontrollierte Beobachtung mit verschiedenen Gruppen durch und konnte so den Effekt verschiedener therapeutischer Maßnahmen direkt miteinander vergleichen. Die systematische Bildung einer Vergleichsgruppe ist das zentrale Element, um den Effekt einer Therapie abzuschätzen. Ohne eine solche Ebene ist es unmöglich, Erkenntnisse über den Wert einer Behandlung zu gewinnen.

Die kritische Auseinandersetzung mit der Wirksamkeit homöopathischer Therapieverfahren führte im 19. Jahrhundert zur Durchführung von Studien, die nach

⁵ Vgl. Breidert et al 2009 S. 754 ff.: Breidert erläutert so zum Beispiel, dass die Wirkung eines Placebos durch statistische Phänomene wie dem Spontanverlauf einer Erkrankung vorgetäuscht werden könnte und dadurch teilweise schlecht abschätzbar sei. Zudem bestünden ethische Probleme.

⁶ Vgl. Lind 1753 Treatise of scuvy.

heutigen Maßstäben einen weiteren Entwicklungsschritt in der klinischen Forschung bedeuten. Die Nürnberger-Kochsalzversuche wurden 1835 durchgeführt.⁷ Ausgangspunkt der Versuche war die Diskussion zwischen Verfechtern und Opponenten der homöopathischen Heilslehre. Im Rahmen dieses Versuchs wurden per Zufall zwei Gruppen gebildet, von denen die eine Fläschchen mit einer homöopathischen Lösung erhielt, die andere Fläschchen gefüllt mit Wasser. Zu welcher Gruppe die Teilnehmer zählten, war weder den Probanden noch den Untersuchern bekannt. Drei Wochen nach Verteilung der Fläschchen wurden die Probanden hinsichtlich der Wirkung befragt und die Ergebnisse wurden ausgewertet.

In diesem Versuch tauchen die Merkmale der *randomisierten Verteilung*, der *Placebokontrolle* und der *doppelten Verblindung* auf. Die randomisierte Verteilung wurde gezielt dazu eingesetzt, eine Selektion zu vermeiden, die durch willkürliche Zuteilung der Probanden entstanden wäre und die Vergleichbarkeit der Gruppen gefährdet hätte. Placebokontrolle und Verblindung wurden in dem Bewusstsein eingesetzt, dass auf diese Art, eine objektive und unvoreingenommene Beurteilung durch Untersucher und Proband möglich ist. Diese Elemente sind mittlerweile prägend für klinische Versuche nach heutigem Standard. Bemerkenswert ist in diesem Zusammenhang die Entstehungsgeschichte der Nürnberger Kochsalzversuche. Diese beruht, wie bereits erwähnt, auf einem Streit zwischen Gegnern und Befürwortern der homöopathischen Heilslehre. Offenbar zielte die penible Versuchsanordnung darauf, unbestreitbare Ergebnisse bezüglich des Nutzens der Homöopathie zu liefern, um diese Kontroverse zu entscheiden.

Etwa zeitgleich zu den Nürnberger Kochsalzversuchen veröffentlicht Louis seine Untersuchungen zum Effekt der Aderlasstherapie.⁸ Bei seinen Untersuchungen bediente er sich der „*numerischen Methode*“. Im Zentrum seiner Methode steht die mathematische Kodierung klinischer Daten wie zum Beispiel der Krankheitsdauer. Nach Sammlung dieser Daten führte er einen statistischen Vergleich von Patientengruppen durch, die zu unterschiedlichen Zeitpunkten mit der Aderlasstherapie behandelt wurden. Auf diese Art gelang es ihm, den Wert der damals gebräuchlichen Aderlasstherapie statistisch zu widerlegen.

Die Einführung einer mathematischen Kodierung stellt eine bedeutende Weiterentwicklung in der klinischen Medizin dar, da diese Aufschlüsselung eine objektive Ergebnisdarstellung ermöglicht. Die Idee einer mathematischen Kodierung stieß zunächst jedoch auf

⁷ Vgl. Löhner 1835

⁸ Vgl. Louis 1835

erheblichen Widerstand seitens klinischer Kollegen von Louis. Das von den Gegnern der numerischen Methode ins Feld geführte Hauptargument richtete sich vor allem gegen die mangelnde Individualisierung des Patienten und des Krankheitsgeschehens.

Interessanterweise begegnet man diesem Argument heute in ähnlicher Form in der Diskussion um die praktische Anwendbarkeit der Evidence based Medicine wieder.⁹

3. Entstehungszusammenhang von Martinis Methodenlehre

Diese Beispiele verdeutlichen, dass die Voraussetzungen, sowohl was die Durchführung des klinischen Versuchs an sich als auch die mathematische Kodierung desselben betrifft, für eine wissenschaftliche Betrachtung in der klinischen Medizin spätestens seit Mitte des 19. Jahrhunderts vorhanden waren. Jedoch scheint die Umsetzung einer kritischen Prüfung in der klinischen Medizin mit wissenschaftlichen Mitteln in der Breite der Forschung in Deutschland kaum stattgefunden zu haben. Ein knappes Jahrhundert nach Veröffentlichung der Nürnberger-Kochsalzversuche und den Untersuchungen von Louis kommentiert Paul Martini, Professor für innere Medizin an der Medizinischen Hochschule in Bonn, den Zustand der klinischen Forschung mit folgenden Worten:

„Durchsucht man ein Dutzend Bände unserer medizinischen Wochenschriften, so hat man schon Glück, wenn man insgesamt 2 Dutzend therapeutische Archive findet, die einer strengen Kritik standhalten. In den klinischen Archiven usw. aber finden sich therapeutische Arbeiten nur verschwindend selten; auf 100 experimentelle oder diagnostische Arbeiten treffen hier vielleicht 3 oder 4 therapeutische (die übrigens zu einem sehr großen Prozentsatz methodologisch ebenfalls nicht genügen). Man geht kaum fehl in der Annahme, dass der Großteil der therapeutischen Autoren sich in einer wissenschaftlich strengeren Umgebung nicht wohl fühlt.“¹⁰

Die Unzufriedenheit mit der methodischen Wertigkeit klinischer Versuche veranlasste Martini 1932 dazu, die „Methodenlehre der therapeutischen Untersuchung“ zu veröffentlichen. Er verfolgte damit das Ziel, der klinischen Forschung ein methodisch solides Fundament zu geben.

In der wissenschaftstheoretischen Forschung wurde die Bedeutung der Methodenlehre Martinis in verschiedenen Zusammenhängen untersucht. So wird ihr ein gewisser Einfluss

⁹ Vgl. auch Tröhler 1993, Elkeles 1999 und Raspe 2003: Louis wurde von den Gegnern der „numerischen Methode“ für eine mangelnde Individualisierung des Patienten kritisiert. Das Argument einer mangelnden Individualisierung des Patienten wird auch in der Diskussion um den Nutzen der Evidence based medicine von den Gegnern dieser aufgegriffen.

¹⁰ P. Martini 1932 S. 39.

für die Einführung der Versuchsverblindung in der klinischen Forschung zugeschrieben.¹¹ Seydel, die die Entwicklung der medizinischen Statistik untersuchte, misst der Leistung Martinis eine herausragende Bedeutung in dem Bereich der Biometrik zu:

„Die Therapeutik, die Methode zur Erkennung einer wirksamen Behandlungsweise, war durch die Statistik begründet worden. Schon vor mehr als 100 Jahren wurde das Für und Wider der Anwendung statistischer Methoden in der Medizin diskutiert. Dieser damals bereits erreichte Wissensstand war jedoch bis zum Erscheinen von Paul Martinis Methodenlehre der therapeutischen Untersuchung 1932 verschüttet.“¹²

Diese Einschätzungen heben einzelne qualitative Merkmale der Methodenlehre hervor. Betrachtet man aber die Methodenlehre in ihrer Gesamtheit unter Berücksichtigung ihres Entstehungszusammenhangs, so wird deutlich, dass sich die Methodenlehre vor allem durch ihr innovatives Konzept auszeichnet und erstmals eine Lücke in der Didaktik der methodischen Durchführung klinischer Versuche schließt, indem sie den klinischen Versuch vollständig von der Planung bis zur Auswertung abbildet.¹³

Die Beschäftigung mit der Methodenlehre diente Stoll als Grundlage für die Überprüfung der Frage, ob es eine deutsche Vorgeschichte der „*Evidence based Medicine*“ gegeben habe. Dabei überprüfte sie die weiter oben beschriebene Kritik Martinis auf der Grundlage einer qualitativen und quantitativen Evaluation klinischer Studien, die zwischen 1881 und 1932 erschienen sind. Stoll bestätigt in ihrer Analyse die Rechtmäßigkeit von Martinis Kritik. Zwar habe es durchaus Beispiele für eine differenzierte methodische Durchführung klinischer Versuche gegeben. Diese Beispiele müssten jedoch eher als Ausnahmen betrachtet werden und könnten nicht darüber hinweg täuschen, dass der Großteil klinischer Versuche in diesem Zeitraum methodisch unpräzise durchgeführt worden sei.¹⁴ Insofern bestand die Notwendigkeit, das Vorgehen klinischer Versuche kritisch zu hinterfragen und der klinischen Forschung ein methodisches Konzept zu vermitteln. In diesem Bereich stellt die Methodenlehre einen zentralen Beitrag dar. Zusammenfassend kommt Stoll zu dem Ergebnis, dass,

„...Martinis Werk einen ganz wesentlichen zusammenfassenden und strukturierenden Beitrag zur Entwicklung der klinisch-therapeutischen Forschung zu Beginn des 20. Jahrhunderts dar(stellt) und mit Sicherheit als ein Grundlagenwerk einer deutschen Vorgeschichte der Evidence based Medicine angesehen und bezeichnet werden (dürfe).“¹⁵

¹¹ Vgl. Winau 1986 S. 100; Vgl. Kaptchuk 1998 S. 421.

¹² Seydel 1976 S. 88.

¹³ Stoll 2005 S. 192.

¹⁴ Stoll 2005 S. 115.

¹⁵ Stoll 2005 S. 192.

Stolls Vorgehen und Ergebnisse dienen als Ausgangspunkt für die vorliegende Arbeit. Dabei steht die Überprüfung der Frage im Vordergrund, wie sich die klinische Forschung in Deutschland nach 1932 methodisch entwickelt hat. Auch für diese Arbeit lässt sich ein Zitat Martinis als zu prüfende Arbeitshypothese anbringen. In der 2. Auflage der Methodenlehre, die 1944 geschrieben, aber erst 1947 veröffentlicht wurde, schreibt Martini: “Seither sind über hundert Jahre vergangen. Dennoch verschafft sich auch heute noch der Großteil der therapeutisch-klinischen Forscher kein zuverlässiges Maß, die beobachteten Vorgänge zu messen, obwohl überall, wo Wert oder Unwert eines Ereignisses bewiesen werden soll, das unbedingte Bedürfnis nach einem Maßstab besteht, mit dem das Ergebnis bzw. seine Folgen verglichen werden können. Es haben sich weder der therapeutische Vergleich noch vergleichende mathematische Operationen auf gesicherter Grundlage durchgesetzt, geschweige denn die Anwendung der Wahrscheinlichkeitsrechnung. Diese aber bilden die unabdingbaren Voraussetzungen der therapeutischen Forschung.“¹⁶ Bereits im Vorwort bewertet Martini den Erfolg der 1. Auflage der Methodenlehre zurückhaltend:

„Nach 13 Jahren weiter Erfahrung hofft das vorliegende Buch diese Lücke vollkommener schließen zu können, als es seinem Vorgänger, der „Methodenlehre der therapeutischen Untersuchung“ 1932 möglich war.“¹⁷

Um Martinis Beurteilung aus der zweiten Auflage von 1947 zu prüfen, wurde die klinisch-therapeutische Forschung unter verschiedenen Aspekten untersucht. Zunächst wurde ein Zeitfenster definiert, in dem die Untersuchung stattfindet. Dabei bestand besonderes Interesse, eine inhaltliche und zeitliche Kontinuität zu Stolls Arbeit herzustellen. Stoll hat sich in ihrer Arbeit mit der Situation der klinischen Forschung zwischen 1881 und 1932 auseinandergesetzt. Folglich beginnt der Untersuchungszeitraum mit dem Jahr 1933. Bei der Wahl des Zeitfensters galt es auch, auf politische Ereignisse Rücksicht zu nehmen. Der zweite Weltkrieg zwischen 1939 und 1945 hat zu einer dramatischen Abnahme von medizinischen Veröffentlichungen geführt. Daher sollte der Endzeitpunkt der Untersuchung nicht unmittelbar mit dem Ende des Krieges übereinstimmen, sondern es musste ein Zeitpunkt gewählt werden, der einen gewissen Abstand zum Jahr 1945 erreicht. Auf der anderen Seite sollte die Entwicklung der klinisch-therapeutischen Forschung vor allem auch in einem engen zeitlichen Zusammenhang mit dem Wirken von P. Martini analysiert werden. Ein zu großes Zeitintervall bis ins Jahr 1960 oder 1970 zum Beispiel

¹⁶ Martini 1947 S. 4.

¹⁷ Martini 1947 im Vorwort.

erschien für diese Fragestellung ungeeignet, da ein Zusammenhang zwischen der Gestaltung klinischer Studien und der erstmals 1932 erschienen Methodenlehre von Martini sich ungenauer hätte verfolgen lassen können. Dies ist vor allem damit zu begründen, dass im angelsächsischen Raum ab 1947 die ersten RCTs durchgeführt worden sind. In den RCTs finden sich wesentliche Merkmale wieder, wie sie Martini für die Durchführung klinischer Versuche beschreibt. Aufgrund dieser Parallelen, zwischen dem, was in der Methodenlehre gefordert wurde und dem, was in den ersten RCTs umgesetzt wurde, erscheint der Versuch, den Effekt der Methodenlehre indirekt nach 1950 zu messen, aufgrund der unklaren Zuordnung nicht sinnvoll.

Im ersten Kapitel sollen zunächst einige therapeutische Studien exemplarisch untersucht werden. Damit werden verschiedene Ziele verfolgt. Zunächst soll ein allgemeiner Eindruck über die damals übliche Publikationspraxis vermittelt werden. Das scheint insofern sinnvoll, als dass es Unterschiede zu der heutigen Publikationspraxis gegeben hat und eine Beurteilung der Studien mit den heute üblichen Maßstäben zu problematischen Schlussfolgerungen führen könnte, lässt man diese Tatsache außer Acht. Ein anderer Zweck, der mit dieser exemplarischen Darstellung verfolgt wird, ist es, zu überprüfen, wie klinische Forschung umgesetzt wurde, welche methodischen Merkmale bekannt waren und zu welchem Zweck diese eingesetzt wurden. Außerdem sollen externe Faktoren und deren Auswirkungen auf die klinische Forschung hinterfragt werden.

Ergänzt wird dieses zunächst rein qualitative Vorgehen durch eine quantitative Analyse klinischer Studien. Im Vordergrund dieses Teils steht die systematische Erfassung und Auswertung klinischer Studien unter methodischen Gesichtspunkten. Dabei gilt ein besonderes Interesse, ob und in welchem Maß die von P. Martini formulierten Vorschläge zur Durchführung klinischer Studien umgesetzt wurden.

Des Weiteren ist eine Darstellung des zeitlichen Verlaufs klinischer Studien unter methodischen Gesichtspunkten von Interesse. Dieses Interesse wird vor allem durch die zu Anfang erwähnte Beurteilung von Martini bestärkt, der keine positive Entwicklung in der klinischen Forschung beobachten kann. Dabei bietet es sich an, einen zeitlichen Verlauf in drei verschiedenen Perioden darzustellen. Die Bildung der Perioden orientiert sich vor allem an externen politischen Ereignissen. Die erste Periode beginnt im Jahr 1933 und endet 1938 und kann auch als Vorkriegsperiode bezeichnet werden. Die Kriegsperiode reicht von 1939 bis 1945 und die Nachkriegsperiode von 1946 bis 1950. So kann ein zeitlicher Verlauf sinnvoll dargestellt werden. Außerdem lässt sich durch Anwendung

einer vergleichbaren Methodik auch eine Verbindung zu den Ergebnissen von Stoll herstellen, so dass eine Entwicklung von 1918 bis 1950 überprüft werden kann. Gezielt soll außerdem untersucht werden, welchen Einfluss die Methodenlehre auf die klinische Forschung gehabt hat. Wie hat sich die Methodenlehre zwischen den ersten beiden Auflagen verändert, welche Inhalte der Methodenlehre wurden wahrgenommen und wie wurden diese unter den klinischen Forschern diskutiert? Das sind die Fragen, die in einem gesonderten Teil der Arbeit beleuchtet werden.

Abschließend soll noch auf zwei weitere Aspekte eingegangen werden. Einer betrifft die Frage nach dem Einfluss ausländischer Forschung auf die deutsche. Diese Frage ist insofern von Interesse, als dass die ersten RCT`s in angelsächsischen Ländern durchgeführt wurden und in Deutschland diese Art der Studiendurchführung scheinbar lange Zeit ignoriert wurde.¹⁸ Insofern soll überprüft werden, ob sich in deutschen Studien Hinweise auf eine Kommunikation mit angelsächsischen aber auch mit anderen Autoren aus dem Ausland auffinden lassen und ob sich Veränderungen in dieser Kommunikation beobachten lassen.

Der zweite und letzte zu untersuchende Aspekt betrifft die Frage nach dem Einfluss der politischen und ethischen Situation auf die klinische Forschung. So finden sich bei verschiedenen Autoren Hinweise darauf, dass die methodische Entwicklung klinischer Studien nach 1945 durch Menschenversuche in Konzentrationslagern oder in militärischen Lagern belastet war.¹⁹ Das wurde vor allem durch die Gleichsetzung von kriminellen Menschenversuchen in Konzentrationslagern und militärischen Einrichtungen mit klinischen Versuchen in Krankenhäusern begründet.²⁰ Insofern soll eine besondere Betrachtung militärischen Studien gewidmet werden unter besonderer Berücksichtigung der ethischen Voraussetzungen zur Durchführung dieser Versuche.

¹⁸ Tröhler 1993 S. 73

¹⁹ Tröhler 1993 S. 73

²⁰ Deutsch 1997 S. 108

2. Material und Methoden

Ziel der vorliegenden Arbeit ist die Erfassung der methodischen Entwicklung der klinisch-therapeutischen Forschung in Deutschland im Zeitraum zwischen 1933 und 1950. Um sich diesem Ziel anzunähern, ist es zunächst notwendig, einzelne Problemfelder und Fragestellungen zu formulieren. Im Einzelnen gliedern diese sich wie folgt:

1. Im ersten Teil der Arbeit wird auf die Situation der klinisch-therapeutischen Forschung in der inneren Medizin zwischen 1933 und 1950 eingegangen. Dazu werden klinisch-therapeutische Studien aus medizinischen Zeitschriften extrahiert und auf bestimmte Merkmale hinsichtlich der methodischen Durchführung untersucht. In diesem Zusammenhang wird auf zwei Fragestellungen im Besonderen Bezug genommen:
 - a. Wie wurde klinische Forschung in Deutschland praktiziert?
 - b. Ergeben sich Unterschiede zu der Zeit vor 1933 und innerhalb des Zeitraums zwischen 1933 und 1950?

Der Bearbeitung der Fragestellung 1 a) werden zwei Teile gewidmet. Im ersten Teil werden Studien, die in dem Zeitraum zwischen 1933 und 1950 erschienen sind, ausgewählt und an deren Beispiel werden methodische Verfahren erläutert. Im zweiten Teil wird eine quantitative Untersuchung des empirischen Materials durchgeführt.

2. Im zweiten Teil wird auf die Rolle von Paul Martini eingegangen. Dabei soll vor allem überprüft werden, in welchem Umfang seine Ideen zur Durchführung klinischer Forschung zur Kenntnis genommen, wie sie umgesetzt und diskutiert wurden.
3. Der dritte Teil beschäftigt sich mit der Frage, ob es externe Einflüsse auf die methodische Entwicklung in Deutschland gegeben hat und wie diese in ihrem Wirkungsgrad einzuschätzen sind. Dabei ist zum einen die Rolle der Forschung aus dem Ausland zu überprüfen. Zum anderen soll auf die klinische Forschung in militärischen und anderen nicht zivilen Einrichtungen eingegangen werden.

1. Untersuchung der klinischen Forschung in Deutschland

Im Zentrum des ersten Teils steht die Frage nach der methodischen Durchführung klinischer Studien und nach möglichen Veränderungen in diesem Bereich. Um dieses Thema zu bearbeiten, war die Auseinandersetzung mit der Arbeit von Stoll von Bedeutung, und daher wird im Folgenden kurz die Methodik von Stoll skizziert und anschließend detailliert auf die Methodik der vorliegenden Arbeit eingegangen.

Stoll hat sich in ihrer Dissertation „Evidence based Medicine“

– Gibt es eine deutsche Vorgeschichte?“²¹ unter anderem mit der Frage nach der Entwicklung der klinisch-therapeutischen Forschung in Deutschland zwischen 1881 und 1932 beschäftigt. Dabei fokussiert sich der Schwerpunkt ihrer Arbeit auf die Zeit zwischen 1918 und 1932. Als Grundlage ihrer Untersuchung diente die Auseinandersetzung mit klinischen Studien aus dem besagten Zeitraum. Die ausgewählten Studien wurden hinsichtlich ihres methodologischen Aufbaus quantitativ und qualitativ erfasst.²²

Die Studien wurden aus medizinischen Zeitschriften identifiziert. Bei der Auswahl der medizinischen Zeitschriften hat Stoll verschiedene Faktoren berücksichtigt. Als die wichtigsten Kriterien zur Berücksichtigung einer medizinischen Zeitschrift für ihre Arbeit galt deren Rezeption und Verbreitung unter deutschen Ärzten. Die Abonnentenzahl wurde als Maßstab für die deutschlandweite Verbreitung und Bedeutung definiert.²³

In den ausgewählten Zeitschriften wurden Studien identifiziert, die einen klinisch-therapeutischen Forschungsschwerpunkt aufweisen. Diese Studien hat Stoll dann nach einer gemeinsam mit Raspe entwickelten Checkliste hinsichtlich des Studiendesigns ausgewertet, auf deren Inhalt ich noch eingehen werde.

In Anlehnung an die Methodik von Stoll habe ich zunächst ebenfalls eine Auswahl von Zeitschriften vorgenommen. Dabei galt es, auf verschiedene Gesichtspunkte Rücksicht zu nehmen. Zum einen bestand das Interesse, die Kontinuität zu der Arbeit von Stoll zu wahren. Zum anderen musste natürlich berücksichtigt werden, dass zwischen 1918 und 1932 medizinische Zeitschriften neu erschienen sind, die Stoll nicht berücksichtigt hat. Um der Forderung nach Kontinuität nachzukommen, wurde primär auf die Zeitschriftenauswahl von Stoll zurückgegriffen. Stoll hat folgende Zeitschriften berücksichtigt:

²¹ Stoll 2005.

²² Vgl. Stoll 2005 S. 16.

²³ Stoll 2005 S. 24.

- Deutsche Medizinische Wochenschrift
- Münchner Medizinische Wochenschrift
- Klinische Wochenschrift²⁴
- Deutsches Archiv für Klinische Medizin
- Verhandlungen der Deutschen Gesellschaft für klinische Medizin

Zwei weitere Zeitschriften wurden von Stoll zunächst mitberücksichtigt, später jedoch aus der Auswertung wegen mangelhaften Bestands ausgeschlossen. Ausschlaggebende Faktoren für die Berücksichtigung bei der Auswahl waren nach Stolls Angaben, zum einen der Bekanntheitsgrad gemessen an der Abonnentenzahl und zum anderen das Publikationsverzeichnis von Paul Martini.

Um Neuauflagen von Zeitschriften zwischen 1918 und 1932 nicht unbeachtet zu lassen, habe ich mich an die Zentralbibliothek in Berlin gewandt, um Informationen über medizinische Zeitschriften zu erhalten, die in dem Zeitraum zwischen 1918 und 1932 neu erschienen sind. Aus einer Liste, in der sämtliche von der Bibliothek geführten medizinischen Zeitschriften aufgelistet sind, ließ sich verfolgen, welche Zeitschriften in dem erwähnten Zeitraum neu erschienen sind und wie lange diese verlegt wurden. Für die Entscheidung, ob diese Zeitschriften ebenfalls als Quelle für die Studien hinzugezogen werden sollten, wurde dann überprüft, ob die Zeitschriften kontinuierlich erschienen sind. Dies ließ sich zunächst an den Angaben der Zentralbibliothek verfolgen. Da die Angaben der Zentralbibliothek über den Bestand allerdings unpräzise sind, was den genauen Erscheinungszeitraum und eventuelle Lücken betrifft, wurden in dem Zeitschriftenmagazin der Hochschulbibliothek in Lübeck die Recherchen darüber fortgeführt und vervollständigt. Als mögliche ergänzende Quellen für die zu untersuchenden Studien wurden die Zeitschriften „Medizinische Klinik“, die ab 1914 regelmäßig erschien und „Die medizinische Welt“, die ab 1927 in den Druck ging, ausfindig gemacht. Obwohl weder beim Verlag noch bei der Zentralbibliothek Angaben über die Auflagenstärke der beiden Zeitschriften ermittelbar waren, kann davon ausgegangen werden, dass es sich um Zeitschriften mit überregionaler Präsenz handelt. Als Hinweis auf die überregionale Bedeutung wurde die Tatsache gewertet, dass diese Zeitschriften an verschiedenen Bibliotheksstandorten verfügbar sind und beide Zeitschriften mit der „Deutschen Medizinischen Wochenschrift“, der „Klinischen Wochenschrift“ und der „Münchner

²⁴ Die Zeitschrift Klinische Wochenschrift ist aus den Vorgängerzeitschriften Berliner Klinische Wochenschrift und Therapeutische Monatshefte hervorgegangen; unter dem Titel Klinische Wochenschrift erscheint sie seit 1922.

Medizinischen Wochenschrift“, 1945 aufgrund des Papier- und Rohstoffmangels zur „Medizinischen Zeitschrift“ zusammengefasst wurden.

Bei der stichprobenartigen Durchsicht der Bände stellte sich heraus, dass in beiden Zeitschriften anstelle von therapeutischen Studien zumeist Therapieempfehlungen ohne wissenschaftlichen Anspruch veröffentlicht wurden. Diese Therapieempfehlungen waren vor allem an den praktizierenden Arzt gerichtet und sollten diesem als Unterstützung bei der Auswahl eines Therapiekonzeptes dienen. Therapeutische Studien im wissenschaftlichen Sinn wurden verglichen mit den anderen medizinischen Zeitschriften wie DMW oder MMW in geringem Maß veröffentlicht. Daher erschien die weitere Berücksichtigung dieser Zeitschriften verzichtbar.

Nachdem die oben erwähnten Zeitschriften ausgewählt worden waren, wurden die für die Fragestellung relevanten Artikel herausgesucht. Die alleinige Zuhilfenahme des Registers und des Inhaltsverzeichnisses, in denen unter Ausdrücken wie „Zur Therapie/Behandlung von...“ oder ähnlichen Formulierungen nach Artikeln über klinisch-therapeutische Studien gesucht wurde, erschien zwar hilfreich aber nicht ausreichend. Die Inhaltsverzeichnisse und Register der Bände waren zum einen unvollständig und zum anderen wurden nicht alle klinisch-therapeutischen Studien über derartige Formulierungen abgebildet. Daher wurden die Bände auf der Suche nach klinisch-therapeutischen Studien Seite für Seite überprüft. Dabei wurden zunächst die Artikel in eine Liste eingetragen, deren Titel oder Inhalt auf eine klinisch-therapeutische Fragestellung hinweist. Wichtig ist es, an dieser Stelle zu betonen, dass in diesem Arbeitsschritt möglichst sensitiv vorgegangen werden musste, um zu vermeiden, dass wertvolles Material der weiteren genaueren Evaluation entgeht.

Daran anschließend wurden die ausgesuchten Artikel einer genaueren Betrachtung unterzogen, in der festgestellt werden sollte, ob in dem ausgesuchten Artikel tatsächlich die Merkmale einer klinisch-therapeutischen Studie beschrieben werden. Zu diesem Zweck wurden die Artikel grob kategorisiert und es wurde festgestellt, ob der Artikel für eine detaillierte Untersuchung infrage kommt. Folgende Kategorien wurden definiert:

„Beobachtungsstudie“, **„Kasuistik“**, **„Diskussion“**, **„Fallserie“**, **„Querschnittsstudie“**, **„Experiment“**.

Artikel, die als Diskussion oder Kasuistik eingeordnet wurden, wurden unberücksichtigt gelassen. Als Diskussion wurden Artikel bezeichnet, die sich vielleicht durchaus mit einer klinisch-therapeutischen Fragestellung auseinandersetzen, in denen aber meist Studien oder Erfahrungsberichte Dritter diskutiert wurden, ohne eine selbst vom Autor durchgeführte Studie zu beschreiben. Diese sind am ehesten mit den heute so genannten

Reviews zu vergleichen. Als Kasuistik wurden die Studien definiert, in denen über die Therapie weniger Probanden oder Patienten mehr oder weniger systematisch berichtet wurde.²⁵ Kasuistische Beobachtungen spielen eine wichtige Rolle für die klinische Forschung, indem diese häufig auf Therapieoptionen aufmerksam machen. Jedoch eignen sie sich für die Bearbeitung dieser Fragestellung nicht.

Die Zugehörigkeit der einzelnen Studien, zu den verschiedenen Kategorien richtet sich nach verschiedenen Gesichtspunkten. Zum einen ist die Anzahl der Probanden oder Patienten ein wichtiger Faktor. Darüber hinaus spielen noch die Bildung von Kontrollgruppen, die Fragestellung der Studie, der zeitliche Beginn des Versuchs, die Auswahl der Versuchsteilnehmer und die Art der Durchführung eine wichtige Rolle für die Entscheidung, in welche der Kategorien sich die verschiedenen Studien einordnen lassen. Als problematisch hat sich die zum Teil sehr ungenaue Beschreibung der Studien herausgestellt. Dadurch war es häufig nicht einfach, eine eindeutige Klassifizierung vorzunehmen, und machte nach Abschluss der primären Auswertung eine Zweitanalyse der ausgesuchten Studien notwendig. Zudem lassen sich heute gültige Definitionen, wodurch zum Beispiel Fallserien, Beobachtungsstudien, Experimente etc. bestimmt sind und wie sie sich voneinander unterscheiden, nicht unbedingt auf das Material anwenden. Das liegt vor allem daran, dass sich allgemein anerkannte Standards der Durchführung klinischer Versuche erst ab Mitte des 20. Jahrhunderts etabliert haben.²⁶ Insofern weichen die in dieser Arbeit verwendeten Definitionen von Studientypen in einigen Bereichen von den heute gültigen ab.

Unter der Kategorie **Experiment** wurden die Studien zusammengefasst, in denen eine bewusste Bildung einer Kontrollgruppe durchgeführt wurde, um Vergleichsbedingungen zu schaffen. Eine wichtige Bedingung für die Kategorisierung als Kontrollgruppe ist, dass aus dem Artikel hervorgeht, dass die Gruppenbildung nicht zufällig und nachträglich, sondern bewusst, vorsätzlich also *interventionell* durchgeführt worden ist. Andernfalls sind diese Studien eher als vergleichende Beobachtungsstudien einzuordnen.²⁷

²⁵ Vgl. Fletcher et al 1999 S. 291; der kasuistische Bericht entspricht der Definition des Fallberichts.

²⁶ Als hinweisend für den geringen Grad der methodischen Standardisierung in der klinischen Forschung lässt sich die Diskussion um die statistische Auswertung klinischer Studien zwischen Martini, Koller und Heidenhain betrachten.

²⁷ Die in dieser Arbeit verwendete Definition des Begriffs „Experiment“ entspricht also der einer „kontrollierten Studie“ (nach Fletcher 1999 S. 193 f.).

Als **Beobachtungsstudien** wurden Studien bestimmt, in denen eine größere Gruppe (≥ 25) von Patienten unter Therapie über einen bestimmten Zeitraum beobachtet wurde. Die Durchführung der Studie ist prospektiv aber auch als historische Kontrolle retrospektiv möglich.²⁸

Einen Sonderfall stellen vergleichende Beobachtungsstudien dar. Darunter sind Studien zu verstehen, in denen zwei verschieden behandelte Gruppen miteinander verglichen wurden, ohne dass dem Vergleich im Unterschied zur experimentellen Studie eine Intervention vorausging.²⁹ Diese Studien können auch als natürliche Kontrollstudien bezeichnet werden.

Unter dem Begriff **Fallserie** sind im Folgenden Studien zu verstehen, in denen eine Gruppe von Patienten (≥ 10) unter Therapie entweder retro- oder prospektiv beobachtet wurde.³⁰ In dieser Arbeit wurde als wesentlicher Unterschied zur Beobachtungsstudie definiert, dass in Fallserien oder Studien die Patienten zuvor bewusst selektiert wurden, während in der Beobachtungsstudie keine Selektion zuvor stattgefunden hat und dadurch alle Patienten in die Auswertung miteinbezogen wurden.³¹ Durch Ungenauigkeiten in der Publikationspraxis bestand bei einigen Veröffentlichungen Verwechslungsgefahr vor allem zwischen Fall- und Beobachtungsstudien.

Unter der Kategorie **Querschnittstudie** oder Prävalenzstudie wurden Studien zusammengefasst, in denen an einer Patientenkohorte eine Untersuchung oder Intervention durchgeführt wurde. Im Unterschied zur Beobachtungsstudie, in der die Patienten im zeitlichen Verlauf also im Längsschnitt erfasst werden, wird in der Querschnittsstudie die Untersuchung oder Intervention zu einem bestimmten Zeitpunkt vorgenommen, ohne dass der zeitliche Verlauf betrachtet wird.³²

Die Artikel dieser oben beschriebenen Kategorien wurden - soweit es sich um klinisch – therapeutische Studien handelt - im abschließenden Schritt anhand der von Stoll und Raspe entwickelten Checkliste analysiert, auf die noch eingegangen wird.

²⁸ Fletcher et al 1999 S. 142.

²⁹ Vgl. dazu auch § 4 Absatz 23 des AMG; demnach liegt eine nichtinterventionelle Prüfung vor, wenn Beobachtungen ohne einen vorab festgelegten Prüfplan im Rahmen der ärztlichen Praxis zustande gekommen sind.

³⁰ Fletcher et al 1999 S. 296 f.

³¹ Fletcher et al 1999 S. 301; demnach ist stellt die Selektion der Probanden den wesentlichen Unterschied zur Kohortenstudie dar.

³² Fletcher et al 1999 S. 107.

Ein Problem, das bei der Auswahl der für die Auswertung relevanten Artikel auftrat, betraf die Unterscheidung zwischen systematischen Studien und unsystematischen Erfahrungsberichten. Während eine Vielzahl von veröffentlichten Artikeln eine Darstellung von unsystematisch aufgebauten Erfahrungsberichten war, kann man nur einen geringen Teil als Studie bezeichnen. Um beispielhaft die Notwendigkeit der Diskriminierung zwischen systematischen Studien und unsystematischen Erfahrungsberichten deutlich zu machen, wird an dieser Stelle ein in der Münchner Medizinischen Wochenschrift im April 1933 erschienener Artikel in Teilen wiedergegeben. „Wir wollten deshalb untersuchen, ob sich das Lanadigin auch zu einer solchen Digitalistherapie mit großen Dosen eignet.

Unser Material an Herzkranken, die wir im letzten Jahr mit Pandigal (Digitalispräparat; Anmerkung des Autors) behandelt haben, können wir in drei Gruppen einteilen. Die erste Gruppe umfasst Kranke mit schwerster Kreislaufdekompensation, bei denen ein sofortiges therapeutisches Eingreifen notwendig war. In der zweiten Gruppe befanden sich mittelschwere Fälle, bei denen die Digitalistherapie durch Bettruhe und Diät vorbereitet werden konnte. In der dritten Gruppe bleiben die leicht Dekompensierten, die ambulant behandelt werden konnten und Kranke mit Arrhythmia absoluta, bei denen ein Übergang in die hämodynamisch ungünstigere tachykarde Form durch Dauerbehandlung mit kleinen Dosen verhindert werden sollte.

Bei den 24 Fällen der 1. Gruppe haben wir zunächst das Glykosid in einer Dosis von 2mal täglich 0,4 mg bis 2mal 0,6 mg intravenös injiziert, um am 2. oder 3. Tag zur peroralen oder rektalen Verordnung überzugehen.

Es hat sich sehr bald gezeigt, dass auch die perorale und rektale Wirkung sehr rasch, häufig schon am selben Tag auftrat, so dass wir die intravenösen Injektionen nur in den dringendsten Fällen anwandten. Bei der intravenösen Anwendung konnten wir auch bei 10tägiger Verabfolgung von 1,2 mg pro Tag keine Intoxikationserscheinungen, bei sehr gutem therapeutischem Effekt beobachten. ³³

In den folgenden Zeilen geht der Autor zusammenfassend auf die Ergebnisse der Therapie in der ersten Gruppe ein, wobei er auf eine objektivierbare Darstellung der Resultate zum Beispiel durch die Angabe von Messwerten verzichtet. Die Art der Behandlung und die Therapieergebnisse der beiden anderen Gruppen werden in ähnlicher Weise dargestellt, ohne dass auf konkrete Ergebnisse – zum Beispiel durch Messung des Blutdrucks oder des Pulses - eingegangen wird. Der Autor kommt abschließend zu dem Urteil,

³³ Dietrich, Schwiegk, 1933 S. 564.

„...dass die auf Grund des Tierexperimentes gezogenen Schlussfolgerungen durch die günstige Erfahrung am Krankenbett bestätigt worden sind.“³⁴

Am Beispiel dieses Artikels, der stellvertretend für eine große Anzahl von Veröffentlichungen therapeutischer Berichte steht, wird deutlich, dass viele Veröffentlichungen sich im Grenzbereich zwischen einer systematischen Datenerhebung im Sinne einer wissenschaftlichen Studie und einem Erfahrungsbericht mit narrativem Charakter befinden. So enthält der Artikel Elemente, die durchaus auf eine systematische Datenerhebung hinweisen, indem er zum Beispiel zu Anfang eine vage Problemstellung formuliert, die anhand der Beobachtung an den Patienten überprüft werden soll. Die erwähnte Bildung der Subgruppen kann ebenfalls als Indiz verstanden werden, dass der Artikel die Durchführung einer systematischen Studie beschreibt. Allerdings bleibt es unklar, ob die Gruppen vor oder nach Durchführung der Behandlung gebildet worden sind. Auch die Kriterien, nach denen die Patienten den Gruppen zugeordnet worden sind, werden nur ungenau beschrieben. Die durch den Autor erwähnte Ergebnisdarstellung beschränkt sich auf ungenaue Formulierungen und ist für eine systematische Datenerhebung ungeeignet, da keine Messgrößen definiert werden, anhand derer der Erfolg der Behandlung gemessen werden kann.

An diesem Artikel wird deutlich, dass es notwendig ist, Kriterien zu definieren, die dabei behilflich sein sollen, systematische Datenerhebungen von narrativen Erfahrungsberichten zu unterscheiden. Dabei betreffen einige Kriterien die beschriebene Durchführung des Therapieversuchs, während andere Kriterien sich auf die Art und Qualität der Veröffentlichung beziehen.

Diese Kriterien wurden in Zusammenarbeit mit Raspe definiert. Für die vom Autor beschriebene Durchführung eines Therapieversuchs wurden folgende diskriminierende Kriterien definiert:

- *„Ex ante“*-Planung: Von wesentlicher Bedeutung ist, dass eine wissenschaftliche Arbeit nicht durch Zufall entstehen kann, sondern dieser eine Planung vorausgehen muss. Es muss also aus dem Text hervorgehen, dass der Autor zum einen seine Erkenntnisintention dargestellt und begründet hat, und zum anderen die Durchführung entsprechend geplant und organisiert wurde.

³⁴ Dietrich, Schwiegk 1933 S. 564.

- Die **Fragestellung** sollte in der Art formuliert sein, dass sich aus ihrer Beantwortung ein verallgemeinerbarer Wissenszuwachs gewinnen lässt und nicht nur ein fallbezogener.
- Das **Design** der Studie sollte auf die Fragestellung angepasst sein.

Für die Beurteilung der Qualität der Veröffentlichung waren folgende Kriterien von Bedeutung:

- Der Text sollte möglichst **präzise Aussagen** über die Art der Studie, über die Patienten oder Probanden und über die durchgeführten Behandlungen enthalten.
- Die **Ergebnisdarstellung** sollte adäquat und nachvollziehbar sein.
- Die Artikel sollten einen **systematischen** Aufbau aufweisen.

Diese sechs Kriterien sind bei der Unterscheidung zwischen Studie und Erfahrungsbericht als Stütze zu verstehen. Die teils sehr unsystematische und lückenhafte Darstellung der durchgeführten Studien ließ häufig nur vermuten, dass einige der erwähnten Kriterien umgesetzt wurden. Daher wurden die Kriterien der „Ex ante“-Planung³⁵ und der des „Designs“ stärker gewichtet als die übrigen. Mit diesem Vorgehen konnte gewährleistet werden, dass kein für die Arbeit wertvolles Material bei der Evaluation unberücksichtigt blieb.

Nachdem mit Hilfe des oben beschriebenen Auswahlverfahrens bestimmt wurde, ob die Studien in die weitere Bearbeitung mit eingehen sollten, wurden die Studien dann durch die bereits erwähnte Checkliste inhaltlich evaluiert. Hintergrund für die Erstellung der Checkliste ist die Absicht, die ausgewählten Studien hinsichtlich der angewandten Methodik zu beurteilen. Die interessierenden Merkmale der Checkliste sind auf der Grundlage von P. Martinis Forderungen an die klinisch-therapeutische Forschung formuliert worden, die in der 1933 erschienenen Methodenlehre veröffentlicht wurden. Die Fragen, die die Checkliste enthält, lassen sich verschiedenen Themengebieten zuordnen. Die ersten Fragen der Checkliste befassen sich mit dem Thema der **Studienvorbereitung und Planung**. Zu diesem Thema sollte geklärt werden, ob ein Stichprobenplan beschrieben wurde, ob Ein- oder Ausschlusskriterien definiert wurden und in welchem Zeitraum, die Studie durchgeführt wurde.

³⁵ Diese Definition der „Ex ante“-Planung sollte aber keineswegs nichtinterventionelle Studien nach § 4 Absatz 23 des AMG von der weiteren Untersuchung ausschließen.

Der folgende Fragenkomplex bezieht sich ausschließlich auf *mehrrarmige Studien*. Interessierende Merkmale in diesem Themenbereich sind die Gruppenbildung, die Kriterien zur Zuweisung zu den verschiedenen Gruppen und die unterschiedliche Behandlung der Teilnehmer der Gruppen.

Ein weiterer Themenbereich befasst sich mit potentiellen *Störgrößen*, die die Studien beeinflussen. Erfasst wurden die Fragen, ob die Einnahme studienunabhängiger *Medikamente* beschrieben wurde und welchem Zweck diese zusätzlich eingenommen Medikamente dienen.

Mit Fragen zur Outcomesoperationalisierung und zur Verblindung der Behandlung und Messung wurden weitere Aspekte der *Studiendurchführung* erfasst und bewertet. Ergänzt wurde die Evaluation durch die Bearbeitung der Fragen nach der *Studienauswertung*.

Von besonderem Interesse waren in diesem Zusammenhang, ob statistische Analysen zum Einsatz gekommen sind, ob bei der Auswertung Subgruppenanalysen durchgeführt und ob Probanden während der Versuche ausgeschieden sind.

Der abschließende Themenbereich beinhaltet Fragen nach der *kritischen Auseinandersetzung* mit der eigenen Studie. Dazu sollte vor allem überprüft werden, inwiefern sich die Autoren mit methodischen und ethischen Aspekten ihrer Arbeit auseinandergesetzt haben und ob sie die eigene Arbeit kritisch hinterfragt haben.

Um eine Grundlage für die Frage nach der *Rezeption* aus dem *Ausland* zu gewinnen, wurde die Checkliste um einen Punkt erweitert, in dem überprüft wurde, ob der Autor ausländische Arbeiten zitiert hat.

Ein grundsätzliches Problem bei der Evaluation stellt die zum Teil sehr unpräzise Darstellung der Artikel dar. Dadurch bedingt ließen sich häufig keine expliziten Aussagen zu den verschiedenen Fragestellungen auffinden. Da aber häufig bestimmte Formulierungen implizit auf eine Vorgehensweise hinweisen, hat es sich als sinnvoll erwiesen, bei der Bearbeitung bestimmter Punkte zwischen expliziten und impliziten Aussagen zu unterscheiden. So wurde in vielen Veröffentlichungen erwähnt, dass bei der Rekrutierung der Patienten Kriterien zu deren Einschluss (zum Beispiel Patientenalter < 40 Jahre) in der Studie existierten. Ausschlusskriterien wurden oft explizit nicht erwähnt, ließen sich aber in vielen Fällen durch Formulierung von Einschlusskriterien implizit ableiten. Diese Differenzierung zwischen expliziten und impliziten Aussagen wurde in der Auswertung berücksichtigt. Die Checkliste befindet sich im Anhang.

Ein weiterer Punkt, der erörtert werden muss, betrifft die medizinischen Disziplinen, auf die sich die Arbeit bezieht. So stehen Studien aus den Bereichen der inneren Medizin im Fokus des Interesses. Es bestand aber die Notwendigkeit nicht ausschließlich Studien aus dem Bereich der inneren Medizin auszuwerten, sondern darüber hinaus auch Studien aus anderen medizinischen Disziplinen einzubeziehen. Der Grund dafür liegt darin, dass eine zu strikte Fokussierung der Arbeit auf Studien aus dem Bereich der inneren Medizin zu einer unnötigen Beschränkung der Beobachtung geführt hätte. Um zu vermeiden, dass wertvolles Material in der Untersuchung unberücksichtigt bleibt, wurden daher auch Studien aus anderen Disziplinen mitberücksichtigt, sofern die Fragestellung primär klinisch–therapeutischer Natur war und sich nicht etwa durch eine fachspezifische therapeutische Methode auszeichnete, wie zum Beispiel durch chirotherapeutische Interventionen im Bereich der Orthopädie.

Aus dem Bereich der Dermatologie/Venerologie wurden 16 Studien ausgewertet, was einem Anteil an der Gesamtzahl der Studien von etwa 10% entspricht. Damit trägt die Dermatologie nach den Fächern der inneren Medizin den größten Teil zu den ausgewerteten Studien bei. Ein Großteil der dermatologischen Studien ist in militärischen Einrichtungen durchgeführt worden mit dem Ziel neue Erkenntnisse zur Bekämpfung der Gonorrhö zu gewinnen, die offensichtlich ein großes Problem für die Wehrfähigkeit der Soldaten darstellte.

Jeweils sechs Studien (4%) wurden aus der Gynäkologie und der Psychiatrie/Neurologie in die Auswertung mit einbezogen. Vier Studien (3%) sind der Disziplin der Chirurgie zuzuordnen, und eine Studie wurde in der Urologie durchgeführt.

Ausgeschlossen wurden sämtliche Studien, in denen eine chirurgische Intervention durchgeführt wurde. Außerdem sind in der vorliegenden Arbeit ebenfalls Studien aus dem Umfeld der Pädiatrie, der Strahlentherapie oder physikalischer Heilverfahren ausgeschlossen worden.

Es sind auch einige Studien mit in die Auswertung einbezogen wurden, die im Ausland entstanden sind. In den Kriegsjahren betrifft das vor allem therapeutische Studien, die von deutschen Medizinern in den Feldkriegslazaretten durchgeführt wurden. Darüber hinaus wurden auch Studien berücksichtigt, die von Medizinern aus dem Ausland durchgeführt wurden und in den oben genannten Zeitschriften veröffentlicht wurden.

Insgesamt wurden Studien von Autoren aus sechs verschiedenen Nationen ausgewertet. So wurde jeweils eine Studie von rumänischen, türkischen, spanischen und japanischen Autoren ausgewertet. Es wurden drei Studien von dänischen und vier Studien von

ungarischen Autoren berücksichtigt. Wegen der kulturellen Nähe und des Anschlusses Österreichs an Deutschland wurden Studien von österreichischen Autoren nicht als ausländische Studien betrachtet. Alle Studien, die auf dem Gebiet des damaligen deutschen Reichs entstanden sind und bei denen kein Hinweis vorlag, dass es sich um ausländische Autoren handelt, wurden dem Inland zugerechnet.

Ein wesentliches Ziel in diesem Teil ist es, nach Veränderungen in der Forschungsgestaltung in dem Zeitraum zwischen 1933 und 1950 zu forschen und diese gegebenenfalls zu beschreiben. Aufgrund der politischen Ereignisse in dem angegebenen Zeitraum bot es sich an, zwischen 3 verschiedenen zeitlichen Perioden zu unterscheiden. Die Jahre zwischen 1933 und 1939 wurden in der ersten Periode als Vorkriegsperiode zusammengefasst, die zweite Periode bezieht sich auf die Zeit zwischen 1939 und 1945 und die dritte Periode beinhaltet die Nachkriegszeit zwischen 1945 und 1950. Diese Untergliederung in drei verschiedene Zeitintervalle erlaubte eine übersichtliche Darstellung von potentiellen Veränderungen in der klinischen Forschung.

Die ausgewerteten Artikel wurden anschließend den drei verschiedenen Perioden zugeordnet. Dabei ist zu beachten, dass sich die Zuteilung zu den Perioden am Datum orientiert, zu dem die Artikel veröffentlicht wurden. Dieses Vorgehen ist notwendig, da zum einen einige Studien über mehr als nur eine Periode durchgeführt worden sind und zum anderen war das Datum, zu dem die Studien begonnen beziehungsweise abgeschlossen wurden, häufig nicht zu ermitteln. Diese Zuteilung ermöglichte es, eine quantitative Verteilung der Studien auf die verschiedenen Perioden darzustellen.

Nachdem die Artikel den verschiedenen Perioden zugeordnet worden sind, konnten auch die Ergebnisse der Evaluation durch die Checkliste den drei Perioden zugeordnet werden. Für eine übersichtliche Auswertung wurden die Punkte der Checkliste zu verschiedenen Themenkomplexen zusammengefasst und der entsprechenden Periode zugeordnet. Veränderungen in der Durchführung von Forschung wurden auf diese Art quantitativ abgebildet.

Da speziell in der Kriegszeit die Zahl der Veröffentlichungen stark nachlässt, ist es sinnvoll eine Verhältniszahl zwischen der Anzahl therapeutischer Studien und der der Gesamtveröffentlichungen zu bilden. Diese Verhältniszahl ist Ausdruck der Dichte oder Konzentration therapeutischer Studien im Gesamtmaterial und dient der Vergleichbarkeit der zeitlichen Perioden miteinander. Die Anzahl der Veröffentlichungen musste aufgrund ungenauer Angaben in den Inhaltsverzeichnissen der Zeitschriften geschätzt werden. Dazu

wurde in 15 zufällig ausgewählten Heften jedes einzelnen Zeitschriftenbandes die Anzahl von Veröffentlichungen bestimmt und es wurde ein durchschnittlicher Wert pro Heft berechnet. Durch Berücksichtigung der Gesamtseitenzahl pro Band, die angegeben ist, konnte dann ein durchschnittlicher Wert für die Gesamtzahl von Veröffentlichungen pro Zeitschrift geschätzt werden. Die geschätzten Werte der einzelnen Zeitschriften pro Jahr wurden addiert, und so konnte für jede Periode ein Durchschnittswert von Publikationen berechnet werden. Die Angaben wurden wegen möglicher Ungenauigkeiten gerundet. Berücksichtigt wurden Veröffentlichungen mit mindestens 50 Zeilen. Ausgeschlossen wurden Anmerkungen, Zuschriften von Lesern und ähnliche. So ergab sich für den Zeitraum zwischen 1933 und 1938 eine durchschnittliche Anzahl von etwa 2000 Publikationen jährlich. Zwischen 1939 und 1945 fällt der Wert auf etwa 1000 Publikationen und für die Nachkriegszeit ist dieser Wert mit ca. 700 anzugeben. Wie bereits erwähnt, hat die Auseinandersetzung mit der Arbeit von Stoll zu der Aufnahme dieser Arbeit angeregt. Die Ähnlichkeit der Methodik beider Arbeiten macht einen Vergleich der Zeiträume zwischen 1918 und 1933 und zwischen 1933 und 1950 möglich. Um einen übersichtlichen und sinnvollen Vergleich beider Arbeiten zu erreichen, wurden mehrere Fragen der Checkliste zu Fragekomplexen zusammengefasst. Diese Fragekomplexe beschäftigen sich mit verschiedenen Themen der Durchführung von Studien und sind in Analogie zu den bereits oben erwähnten Fragekomplexen definiert worden.

Der erste Komplex geht auf die **Planung** und **Vorbereitung** der Studien ein. In diesem Komplex wurden Fragen zur **Patientenrekrutierung** zusammengefasst.

Der darauf folgende Komplex setzt sich mit den **mehrarmigen** Studien auseinander. Die Fragen nach dem Modus der Gruppenbildung und nach der Gleich- oder Ungleichbehandlung der Gruppenarme wurden in diesem Komplex miteinander verglichen. **Störgrößen** wurden in einem weiteren Fragenkomplex thematisiert, wobei dieser Fragen zur studienunabhängigen Medikamenteneinnahme enthält.

Die letzten beiden Komplexe beschäftigen sich mit dem Vergleich der Themen der **Studienauswertung** und der **kritischen Auseinandersetzung** der Studien.

2. Zur Wahrnehmung von Paul Martini und seiner Methodenlehre

Martini stellt mit der Methodenlehre einen Leitfaden zur methodischen Durchführung klinischer Studien zur Verfügung.³⁶ Motiviert zum Schreiben der Methodenlehre wurde Martini vor allem durch seine Beobachtung, dass nur wenige therapeutische Studien aus seiner Zeit methodischen Ansprüchen genügten.³⁷

Nach heutiger Auffassung ist die 1. Auflage der Methodenlehre von 1932 mit ihren Vorschlägen zur Forschungsgestaltung im zeitlichen Entstehungszusammenhang als bemerkenswerter didaktischer Ansatz im Bereich der Forschungstheorie aufzufassen.³⁸ Dabei sind nicht unbedingt die einzelnen Empfehlungen von Martini als neuartig zu verstehen, denn viele von diesen waren bereits zuvor bekannt. Vielmehr zeichnet sich die Methodenlehre durch ihre Konzeption aus, indem sie Ideen von der Vorbereitung über die Durchführung bis zur Auswertung des klinischen Versuchs in einem Buch komprimiert. Dadurch war sie einerseits kompakt, bildete aber dennoch den klinischen Versuch in seiner Gesamtheit ab und hatte dadurch das Potential, als Leitfaden für den klinischen Versuch allgemein akzeptiert zu werden.³⁹ Aufgrund dieser Eigenschaften wurde im zweiten Kapitel die Frage überprüft, in welchem Umfang Martini beziehungsweise die Methodenlehre mit seinen Empfehlungen wahrgenommen beziehungsweise akzeptiert wurde und wie diese in der klinischen Forschung impliziert wurden.

Bevor auf die Wahrnehmung von Martini in dem Material eingegangen wurde, wurden zunächst die zentralen Forderungen Martinis erläutert. Dabei galt es zu beachten, dass die Methodenlehre in insgesamt vier verschiedenen Auflagen gedruckt wurde. Die erste Auflage ist 1932 unter dem Titel „Methodenlehre der therapeutischen Untersuchung“ erschienen. Die zweite Auflage wurde 1947 unter einem anderen Titel „Methodenlehre der therapeutisch-klinischen Forschung“ publiziert, die dritte 1953 und die vierte 1968. Die erste und zweite Auflage der Methodenlehre sind vor beziehungsweise in dem untersuchten Zeitraum erschienen. Damit stellen die in diesen Auflagen enthaltenen Empfehlungen die Grundlage für die weitere Untersuchung dar, da diese den zeitgenössischen Forschern zugänglich gewesen waren. Aus der Auseinandersetzung mit den beiden ersten Auflagen der Methodenlehre ließen sich außerdem Informationen ihrer Entstehungsgeschichte gewinnen. In diesem Zusammenhang wurde vor allem kurz auf die Bedeutung der psychiatrischen Kollegen Eugen Bleuler⁴⁰ und von Julius Wagner-Jauregg⁴¹

³⁶ Vgl. Stoll 2005 S. 192.

³⁷ Vgl. Martini 1932.

³⁸ Vgl. Winau 1984 S. 100 f., Tröhler 1993 S. 72 f.

³⁹ Vgl. Stoll 2005 S. 192, Kaptchuk 1998 S. 421.

⁴⁰ Martini 1932 S. 39; Eugen Bleuler Schweizer Psychiater (1857 – 1939).

auf die Methodenlehre eingegangen. Zudem bot sich die Gelegenheit, die beiden ersten Auflagen der Methodenlehre gegenüberzustellen und miteinander zu vergleichen.

Nach Darstellung der wesentlichen Forderungen Martinis wurde in den knapp 160 Studien überprüft, inwiefern Martini wahrgenommen wurde. Die Publikationen wurden darauf untersucht, ob Martini beziehungsweise die Methodenlehre in dem Text oder in dem Literaturnachweis erwähnt wurden. Wurde festgestellt, dass Martini oder die Methodenlehre namentlich in dem Text beschrieben wurden, wurde im nächsten Schritt überprüft, welche Inhalte wahrgenommen wurden. In einigen Fällen wurde nicht explizit beschrieben, worauf sich die Autoren bezogen. Daher wurde weiter zwischen ex- und impliziten Aussagen unterschieden.

In einigen Artikeln wurden Anregungen, wie von Martini vorgeschlagen, umgesetzt oder es ließen sich durch den Gebrauch von in der Methodenlehre spezifisch verwendeten Termini Verbindungen zu ihm oder der Methodenlehre vermuten, ohne dass Martini oder die Methodenlehre explizit erwähnt wurden. Zwar sollte der Gebrauch von diesen Begriffen oder die Beschreibung des von ihm vorgeschlagenen Vorgehens nicht als Beweis für einen tatsächlichen Einfluss von Martini missverstanden werden. Dennoch stellt es ein indirektes Indiz für einen Zusammenhang mit den Empfehlungen der Methodenlehre oder ihr verwandten Vorschlägen dar und wurde entsprechend gewertet.

Das zweite Kapitel wird abgeschlossen mit einem in der Klinischen Wochenschrift veröffentlichten Meinungs austausch zwischen Paul Martini, Siegfried Koller,⁴² einem Statistiker und Sozialmediziner und Heidenhain, einem chirurgischen Kollegen.

Thematisiert wurden in dieser Diskussion die in der Methodenlehre gemachten Vorschläge zur statistischen Auswertung klinischer Versuche. Aus der Diskussion ließen sich interessante Hinweise über die Akzeptanz von Martinis Ideen gewinnen.

⁴¹ Martini 1932 S. 4, S. 39; Julius Wagner-Jauregg wird von Martini als Namensgeber für die „Simultanmethode“ – einer alternierenden Methode bei der Kontrollgruppenbildung – bezeichnet; Julius Wagner-Jauregg (1857 – 1940) österreichischer Psychiater und Nobelpreisträger 1927.

⁴² Siegfried Koller (1908 – 1998) deutscher Sozialmediziner und Statistiker.

3. Untersuchung externer Einflüsse

Der dritte Teil ist der Untersuchung von externen Einflüssen auf die klinische Forschung in Deutschland gewidmet. Dabei steht die Bedeutung von zwei Faktoren im Fokus dieser Beobachtung. Zum einen sollte überprüft werden, ob und in welchem Maß Forschung aus dem Ausland in den untersuchten klinischen Studien wahrgenommen wurde. Zum anderen galt es den Einfluss einer militärischen, nicht zivilen Umgebung auf die Durchführung klinischer Versuche zu untersuchen.

Unter den 158 ausgewerteten Studien befinden sich elf Veröffentlichungen, die im Ausland durchgeführt worden sind. Daran und an der Beobachtung, dass in einem großen Teil der Veröffentlichungen ausländische Autoren zitiert wurden, macht deutlich, dass die methodische Entwicklung in der klinischen Forschung auch von Einflüssen aus dem Ausland abhängig gewesen ist und keinesfalls als isoliert zu betrachten ist. Es lässt sich vermuten, dass das Interesse an der Literatur primär inhaltlichen Aspekten gegolten hat. Aber es ist wahrscheinlich, dass auch die methodische Durchführung von Studien wahrgenommen worden ist. Insofern bestand das Interesse, in dem Material zu überprüfen, in welchem Umfang Literatur aus dem Ausland wahrgenommen wurde und ob sich zum Beispiel bei Ausbruch des Krieges Zäsuren in dem Austausch wissenschaftlicher Arbeiten zeigten.

Um einen quantitativen Eindruck über die Rezeption aus dem Ausland zu erhalten, wurde die von Stoll entwickelte Checkliste um eine Kategorie ergänzt, mit der aufgeklärt werden sollte, ob der Autor der untersuchten Studie sich mit Literatur aus dem Ausland beschäftigt hat. Dabei hat sich herausgestellt, dass das zunächst favorisierte Vorgehen, einzelne Staaten oder Nationen von zitierten Autoren anzugeben, nicht zweckmäßig ist, da sich zum einen eine genaue Staatszugehörigkeit bei angelsächsischen Autoren aufgrund mangelhafter Angaben im Artikel häufig nicht bestimmen ließ, zum anderen für bestimmte Nationen nur sehr geringe Zahlen zustande gekommen wären. So wurden Autoren zitiert, deren Namen – zum Beispiel John Scott – den starken Verdacht nahe legen, dass eine angelsächsische Staatsbürgerschaft vorliegt, ohne dass Aussagen über die genaue Staatszugehörigkeit des Betreffenden gemacht werden konnten, da der Artikel nur unzureichend darüber informiert. Daher ließ sich häufig nicht diskriminieren, ob – wie in dem Beispiel angeführt – die betreffende Person englischer, irischer oder vielleicht US-amerikanischer Staatsbürger gewesen ist.

Um eine übersichtliche und sinnvolle Darstellung zu bieten, habe ich mich dann dazu entschieden, für diese Kategorie die Antwortmöglichkeiten „angelsächsisches Ausland“,

„sonstiges Ausland“ und „keine Angabe“ zu definieren. Dass die Antwortmöglichkeit „sonstiges Ausland“ gegenüber der Antwortmöglichkeit „angelsächsisches Ausland“ relativ grob und ungenau gewählt ist, ist Folge der Annahme, dass die angelsächsischen Länder als politisch unabhängiger Raum im Zeitraum zwischen 1933 und 1950 die größte Bedeutung für die wissenschaftliche Forschung und ihre methodische Entwicklung gehabt haben.⁴³ Für Frankreich und Dänemark zum Beispiel, die 1940 ihre politische Souveränität verloren haben, muss davon ausgegangen werden, dass diese Länder auch auf dem Gebiet der Wissenschaft in ihrer Unabhängigkeit stark eingeschränkt gewesen waren und von 1940 bis kurz vor Kriegsende unter dem Einfluss des Deutschen Reiches gestanden haben. Daher habe ich als Schwerpunkt dieser Betrachtung vor allem die Wahrnehmung aus dem angelsächsischen Sprachraum gewählt. Aus dieser Überlegung heraus erklärt sich außerdem auch das gewählte Vorgehen, Artikel, in denen Autoren sowohl aus dem angelsächsischen als auch aus dem „sonstigen“ Ausland zitiert wurden, nur der Antwortmöglichkeit „angelsächsisches Ausland“ zuzuordnen und diese nicht etwa in beiden Kategorien aufzuführen. Bei der Zuordnung der Studien zu den Antwortmöglichkeiten wurden auch indirekte Hinweise auf die Herkunft der zitierten Autoren als Entscheidungskriterium verwendet. Zwischen impliziten und expliziten Angaben wurde nicht weiter differenziert. Da anhand dieses Vorgehens allerdings nur eine vage Annäherung an die Frage nach dem Einfluss des Auslands auf die klinische Forschung in Deutschland möglich ist, wurde ergänzend eine bemerkenswerte Studie aus Dänemark detailliert analysiert und vorgestellt. Diese ist 1938 also vor der Besetzung Dänemarks durch das Deutsche Reich in der Klinischen Wochenschrift veröffentlicht worden und zeichnet sich durch eine außergewöhnliche methodische Durchführung aus, anhand der sich wertvolle Informationen über den methodischen Stand in Dänemark gewinnen lassen.

Im zweiten Teil dieses Kapitels wurde auf die Situation der klinischen Forschung in militärischen und nicht zivilen Einrichtungen eingegangen. Spätestens ab Kriegsbeginn 1939 spielte die Lazarettforschung eine zunehmende Bedeutung für die klinische Forschung insgesamt. Wurden in der Zeit zwischen 1933 und 1939 6 % der ausgewerteten Studien in militärischen Einrichtungen durchgeführt, sind es während des Krieges 24 %. Die Voraussetzungen für die Durchführung therapeutischer Studien in militärischen oder nicht zivilen Einrichtungen unterscheiden sich von denen in zivilen. Diese Unterschiede

⁴³ Tröhler 1993 S. 73.

reichen von logistischen über inhaltliche bis zu ethischen Aspekten.⁴⁴ Ziel ist es, die zu der zivilen Forschung unterschiedlichen Voraussetzungen quantitativ und qualitativ darzustellen.

Bei der Bearbeitung dieser Aufgabe war es zunächst wichtig, Versuche aus einem militärischen Kontext zu identifizieren. Dazu wurde in den Überschriften der Veröffentlichungen und in den Texten nach Begriffen wie „*Stabs*“- oder „*Oberstabsarzt*“, „*Lager*“- oder „*Lazarettarzt*“ oder ähnlichen gesucht. Aus anderen Veröffentlichungen ging nicht klar hervor, in was für einer Art von Einrichtungen die Versuche durchgeführt wurden. Das war vor allem dann der Fall, wenn die Autoren keine militärische Zusatzbezeichnung aufführten oder aber im Rahmen einer Kooperation zwischen zivilen und militärischen Einrichtungen Versuche durchgeführt wurden. Das konnte nicht immer klar erfasst werden und daher ist es möglich, dass einige Studien fälschlicherweise dem zivilen Bereich zugeordnet wurden. Nachdem die Studien entsprechend identifiziert wurden, wurden diese zu einer Gruppe zusammengefasst und isoliert auf der Grundlage der Checkliste evaluiert. Dadurch ließ sich ein quantitativer Eindruck gewinnen, wie in den Lazaretten und Lagern methodisch vorgegangen worden ist. Ergänzend wurden Studien exemplarisch dargestellt, um auf Besonderheiten in der Planung, Durchführung und im Umgang mit den Versuchsteilnehmern aufmerksam zu machen.

⁴⁴ Vgl. Schmiedebach 1984 S. 193 ff.: Schmiedebach beschreibt die Aufgabe des Arztes in der nationalsozialistischen Gesellschaft in der Wiederherstellung und dem Erhalt der Gesundheit des „Volkskörpers“, die individuelle Gesundheit spielte insofern nur eine untergeordnete Rolle.

3.1. Darstellung der klinischen Forschung zwischen 1933 und 1950 auf der Grundlage exemplarischer Studien.

Zu Anfang des Ergebnisteils soll an dem Beispiel verschiedener Artikel aus dem Zeitraum zwischen 1933 – 1950 ein Eindruck von dem Stand und der zeitgerechten Qualität der klinisch-therapeutischen Forschung vermittelt werden. Das ist vor allem aus einem Grund nötig. Misst man Studien aus diesem Zeitraum an heutigen Maßstäben, zeigt sich das Problem, dass man dem damaligen Entwicklungsstand der Methodik bei der Beurteilung der Studien nicht gerecht wird. Daher werden bei der exemplarischen Darstellung der Studien zwei wesentliche Ziele verfolgt. Zum einen sollen wichtige methodische Merkmale aufgezeigt und erörtert werden. Zum anderen wird untersucht, mit welcher Intention die Forscher die angewandte Methodik eingesetzt haben.

Die Artikel werden gegliedert nach thematischen Gesichtspunkten beschrieben.

I. Artikel: Das alternierende Verfahren

Der erste zu referierende Artikel wurde von Nissen in der Deutschen Medizinischen Wochenschrift veröffentlicht.⁴⁵

Nissen hat eine experimentelle Studie mit einem für damalige Verhältnisse fortschrittlichen Studiendesign durchgeführt. Es wurden zur Frage der Wirksamkeit der Serumtherapie bei Lungenentzündungen zwei Gruppen gebildet, von denen die eine als Kontrollgruppe gegenüber der Verumgruppe diente. Die Kontrollgruppe und die Verumgruppe sind durch ein *alternierendes Verfahren* gebildet worden:

„Die Therapie wurde, um einen überzeugenden Beweis für den heilenden Effekt zu erhalten, in den ersten Monaten in folgender Weise ausgeführt: Erstens wurden, um eine klinische Auswahl von Anfang an zu vermeiden, sowohl Kranke mit kruppöser als auch solche mit Bronchopneumonie behandelt, und zwar immer sofort nach Aufnahme in die Klinik, also ohne die Sputumuntersuchung und Typendifferenzierung abzuwarten, weil wir damals mit der Neufeldschen Kapselschwellung als Schnellmethode zur Typenbestimmung noch keine Erfahrung hatten. Dann wurde die Behandlung zeitmäßig alternierend festgelegt, indem man nur die an bestimmten Wochentagen eingebrachten Kranken mit Serum behandelte, die übrigen an anderen Tagen eingelieferten dagegen nicht.“⁴⁶

⁴⁵ Nissen 1938 S. 221 – 225.

⁴⁶ Nissen 1938 S. 222.

Nissen beschreibt also ein alternierendes Verfahren und nennt auch den Grund für sein Vorgehen, um nämlich:

„1. eine zuverlässige Kontrolluntersuchung, 2. eine frühzeitige Behandlung, in der Regel am 2.-4. Krankheitstag, 3. ein Vergleich zwischen typenspezifischer und typenunspezifischer Serumtherapie und 4. in der Bronchopneumoniegruppe Kontrollfälle zur Beurteilung der Serumtherapie der reinen Pneumokokkeninfektion.“⁴⁷ zu erhalten.

Nissen betont, dass die Patienten gleichartig behandelt wurden:

„Es wurde nur konzentriertes Pneumokokkenserum und zwar nur Typus I gegeben, Bourroughs Welcomes oder Lederles Präparat, Dosis 40000 Einheiten gleich bei der Aufnahme und 20000 Einheiten nach 12 – 14 Stunden, also im ganzen 60000 Einheiten in etwa 20 -30 ccm Serum in den ersten 24 Stunden.“⁴⁸

Das alternierende Experiment mit zwei Gruppen wurde nach 18 Monaten abgebrochen. Der Abbruch des Experiments sei aus der Sicht von Nissen nötig gewesen, da sich in der Studie die überlegene Wirksamkeit der Serumtherapie gegenüber der Kontrollgruppe gezeigt habe. Daraufhin wurden alle Patienten mit Bronchopneumonie und kruppöser Pneumonie mit der Serumtherapie behandelt.

Das in diesem Artikel erwähnte alternierende Verfahren verdient eine genauere Betrachtung. Die alternierende Methode, durch die zwei verschieden behandelte Gruppen zustande kommen, ist eine für diese Zeit durchaus moderne und gebräuchliche. Der Autor beschreibt, dass die Zuweisung zu einer der Gruppen von dem Datum, an welchem der Patient in die Klinik eingewiesen wurde, abhing. Viele andere zeitgenössische Autoren beschreiben ebenfalls ein alternierendes Verfahren zur Bildung einer Kontrollgruppe. In anderen ausgewerteten Veröffentlichungen, in denen ebenfalls eine alternierende Methode zur Gruppenbildung beschrieben wird, finden sich noch andere Kriterien, die zur Zuweisung in die Verum- oder Kontrollgruppe berücksichtigt wurden. So wird in anderen Studien erwähnt, dass die Zuweisung von der Station, auf der die Patienten lagen, oder von der alphabetischen Reihenfolge der Namen abhing. In den meisten Studien mit einer alternierenden Gruppenbildung ist die Zuweisung allerdings von dem Termin oder der Folge des Erscheinens der Patienten abhängig.

Als Gründe für die Bildung einer Kontrollgruppe nennt Nissen, dass er dadurch die Möglichkeit habe, die Serumtherapie im Vergleich zur Kontrollgruppe zu bewerten.⁴⁹ Er ist sich also darüber bewusst, dass zur Beurteilung der Wirksamkeit einer bestimmten

⁴⁷ Nissen 1938 S. 222.

⁴⁸ Nissen 1938 S. 222.

⁴⁹ Vgl. Nissen 1938 S. 222.

Therapie eine Vergleichskontrolle nötig ist. Diese Feststellung mag heute banal und selbstverständlich erscheinen. Man muss jedoch beachten, dass in vielen Studien aus diesem Zeitraum auf die Bildung von Kontrollgruppen verzichtet wurde, so dass die Einsicht, dass eine Vergleichsgruppe zur Beurteilung der Wirksamkeit notwendig ist, als keinesfalls selbstverständlich zu betrachten ist. Trotz des Mangels einer Kontrollgruppe in vielen anderen Studien oder therapeutischen Beobachtungen wurden oft genug unzulässige Interpretationen der Studienergebnisse hinsichtlich der Wirksamkeit von Medikamenten gemacht. Das vergleichende Experiment mit einer Kontrollgruppe wird zwar durchaus schon von einigen klinischen Forschern praktiziert, viele andere verzichten aber darauf und sind sich der Notwendigkeit zur Bildung einer Kontrollgruppe nicht bewusst. In dem Artikel finden sich einige Bemerkungen, die auf die Sensibilität Nissens für eine methodisch zuverlässige Durchführung klinischer Forschung hinweisen.

Nissens Erwähnung, dass die Patienten gleichartig, das heißt mit dem gleichen Wirkstoff und der gleichen Dosis behandelt wurden, unterstreicht sein Bewusstsein für methodische Probleme. In den meisten Studien aus dem vergleichbaren Zeitraum finden sich selten derart differenzierte Ausführungen zur methodischen Durchführung klinischer Forschung.⁵⁰

Ein weiterer erwähnenswerter Aspekt dieser Studie stellt die Begründung dar, unter der der Versuch abgebrochen wird. Nissen erwähnt, dass:

„...das Resultat dieser (Anmerkung: alternierenden) Periode uns veranlasst hat, die alternierende Therapie aufzugeben.“⁵¹ Er schließt aus den Beobachtungen der 18 Monate der alternierenden Behandlung, dass die von ihm beobachteten Ergebnisse seiner Ansicht nach hinreichend positiv für die Serumtherapie ausgefallen seien, woraufhin die alternierende Periode beendet wurde und sämtliche Patienten mit der Serumtherapie behandelt wurden.

Der Autor begründet seinen Entschluss, die alternierende Periode abubrechen, nicht näher. Es liegt aber die Vermutung nahe, dass die Gründe statistischer aber auch medizinisch-ethischer Natur waren. Von anderen Autoren, die einen ähnlichen Entschluss gefällt haben, wird häufig erwähnt, dass sie es für nicht vertretbar hielten, eine offensichtlich überlegene Therapieform Patienten aus der Kontrollgruppe vorzuenthalten und dadurch einen abwendbaren Schaden billigend in Kauf nehmen. So lehnt zum Beispiel ein anderer Kollege von Nissen namens Stangemann in der Veröffentlichung „Zur Frage der

⁵⁰ Vgl. Nissen 1938 S. 222.

⁵¹ Nissen 1938 S. 222.

Pneumoniebehandlung mit Eubasinum und dessen Nebenwirkungen“ eine vergleichende Behandlung mit folgender Begründung grundsätzlich ab:

„Wegen der im Schrifttum mitgeteilten guten Erfahrungen mit diesem Mittel (Anmerkung: Eubasinum, C. Wolkewitz) und der weniger günstigen Ergebnisse der beiden letzten Jahre mit Chinin-Kalzium, wurde von einer alternierenden Behandlung abgesehen.“⁵²

Der Entschluss zum Abbruch einer kontrollierten Studie scheint bei eindeutiger Überlegenheit eines Mittels vor diesem Hintergrund durchaus gerechtfertigt zu sein oder macht den Abbruch ethisch eventuell sogar erforderlich. Dennoch müssen die tatsächlichen Beweggründe, die zum Abbruch einer kontrollierten Studie führen, kritisch hinterfragt werden.

In den meisten Fällen kann man davon ausgehen, dass der Arzt aus Sorge um das Wohlergehen des Patienten eine solche Entscheidung getroffen hat.

Unabhängig vom Motiv, das zum Abbruch der Studien führt, ist doch zumindest das Ergebnis der Untersuchung mit Problemen belastet. Eine vorschnelle, positive Bewertung eines Medikaments, die zum Studienabbruch führt, erschwert es, die tatsächliche Wirksamkeit des Medikamentes über einen längeren Zeitraum an einer größeren Patientenzahl zu untersuchen. Dadurch wird die Aussagekraft einer Studie belastet und führt zu zweifelhaften Ergebnissen.

Ob sich der Autor der beschriebenen Studie dieser Problematik tatsächlich bewusst ist, bleibt offen und wird von diesem nicht näher diskutiert. Der Beobachtungszeitraum der alternierenden Periode erstreckt sich über wenigstens 18 Monate, was eine verhältnismäßig lange Zeit darstellt und was man zumindest als Hinweis auffassen kann, dass keine allzu voreiligen Schlüsse hinsichtlich der Wirksamkeit der Serumtherapie getroffen wurden. Was aus der Veröffentlichung der Studie nicht hervorgeht, ist die Frage nach der Behandlung der Kontrollgruppe. Ob die Kontrollgruppe tatsächlich keine Medikamente erhalten hat oder ob sie eine alternative Therapie bekommen hat, wird von dem Autor nicht näher beschrieben. Man kann allerdings davon ausgehen, dass eine alternative Therapie durchgeführt wurde, ohne dass darauf explizit verwiesen wurde.

⁵² Stangemann S. 984.

II. Artikel: Nichtinterventionelle Prüfung

In dem Artikel von Nissen wurde ein alternierendes Verfahren, angewandt. Kriterium zur Zugehörigkeit zu der einen oder anderen Gruppe war der Termin, an dem die Patienten ins Krankenhaus eingewiesen wurden. Wie bereits erwähnt, wurden in anderen Studien andere Kriterien zur Zuweisung in die Verum- oder Kontrollgruppe verwendet. Ein ungewöhnliches und sicher nicht freiwillig gewähltes Kriterium zur Bildung der Gruppen wird in einem Artikel von Dennig und Hoeffler beschrieben. Dieses unfreiwillig gewählte Kriterium ist auf den Umstand zurückzuführen, dass aus einem Mangel an verfügbaren Medikamenten ein alternierendes Verfahren gewählt wurde, das zur Bildung einer Kontroll- und einer Verumgruppe führte. Der Artikel ist 1933 unter dem Titel „Zur Behandlung der Poliomyelitis mit Rekonvaleszentenserum“ in der Münchner Medizinischen Wochenschrift erschienen. In dem Artikel beschreibt Dennig die Versuchsanordnung:

„Zu jener Zeit war in ganz Deutschland kein Rekonvaleszentenserum zu erhalten. Wir haben daher selber Serum gewonnen zuerst von solchen Leuten, die vor einigen Jahren die Krankheit überstanden hatten, dann aber (die größere Menge) von unseren eigenen Kranken nach der 4. Krankheitswoche. Bei dem großen Ansturm reichten die Mengen aber nicht aus, wir konnten nur einen Teil unserer Kranken mit Serum behandeln; die Auswahl erfolgte nicht nach der Schwere der Krankheit, sondern nur nach Maßgabe des vorhandenen Serums.“⁵³

Zunächst muss erläutert werden, ob es sich überhaupt um eine „Studie“ im engeren Sinn handelt. In dem Abschnitt Material und Methoden wurde auf die Schwierigkeit der Differenzierung zwischen Studien und Erfahrungsberichten auf der Grundlage relativ unsystematischer Beobachtungen hingewiesen. Dabei wurde festgestellt, dass eine experimentelle Studie im Wesentlichen eine „ex ante“-Planung voraussetzt. Man muss am ehesten davon ausgehen, dass es sich bei dem oben beschriebenen Versuchsaufbau nicht um ein „ex ante“ geplantes Experiment handelt, da Dennig eher durch äußere Umstände in diesem Beispiel also durch den Mangel an verfügbarem Serum zu einer vergleichenden Beobachtung gezwungen wird. Man muss vermuten, dass Dennig sich bei entsprechender Verfügbarkeit des Serums nicht zur Durchführung eines solchen Versuchs entschlossen hätte. Obwohl sich Dennig unfreiwillig zu dieser vergleichenden Beobachtung gezwungen sah, ist doch bemerkenswert, dass er sich offensichtlich über den Wert seiner vergleichenden Beobachtung im Klaren ist, wenn er schreibt:

⁵³ Dennig 1933 S. 1367.

„So können wir den Krankheitsverlauf mit und ohne Serum innerhalb derselben Epidemie vergleichen. Dies ist überhaupt die einzige Möglichkeit, um zu einem Urteil zu kommen, da ein Vergleich innerhalb verschiedener Epidemien und Gegenden bei dem wechselnden Genius epidemicus nicht statthaft ist.“⁵⁴.

Durch diese Äußerung bringt er sein Verständnis zum Ausdruck, dass zum einen die Notwendigkeit besteht, eine Vergleichsgrundlage zur Beurteilung der Wirksamkeit eines Medikamentes zu bilden und dass zum anderen die Homogenität des Patientenkollektivs eine große Rolle bei der Erfolgsbeurteilung spielt.

Dieser Artikel ist besonders unter dem Aspekt der Unfreiwilligkeit erwähnenswert, weil er ein treffendes Beispiel für die ungewollte Anwendung einer innovativen Methodik liefert. Ob man in diesem Artikel letztendlich von einer „ex ante“ geplanten experimentellen Studie ausgehen kann, ist fraglich. Am ehesten ist die in dem Artikel beschriebene Art der Arzneimittellprüfung vergleichbar mit der Definition der „nichtinterventionellen Prüfung“ von Arzneien aus dem Arzneimittelgesetz.⁵⁵ Hierbei unterscheidet sich die nichtinterventionelle Prüfung von der interventionellen vor allem dadurch, dass kein vorab definierter Prüfplan existiert und dass die Ergebnisse einer solchen Prüfung aus den Beobachtungen aus der ärztlichen Praxis resultieren.⁵⁶

III. Artikel: Verblindung

In einem anderen Artikel, der in der Deutschen Medizinischen Wochenschrift veröffentlicht wurde, beschreibt der Leipziger Professor Adolf Bingel in dem Artikel „Wirkt das Diphtherieheilserum bei der menschlichen Diphtheriekrankheit spezifisch durch seinen Antitoxingehalt oder unspezifisch?“ eine Versuchsanordnung, die sich auf eine bereits 1942 durchgeführte klinische Studie bezieht. Bevor er auf die 1942 durchgeführte Studie eingeht, erläutert er seine Erfahrung, die er durch bereits 1918 durchgeführte Versuche mit ähnlicher Fragestellung gewonnen hat. Dabei beschreibt er sein Vorgehen wie folgt:

„Ich war so vorgegangen, dass genau in der Reihenfolge der Aufnahmen jeder erste Fall Antitoxin, jeder zweite Leerserum bekam, ohne Rücksicht auf das Alter des Kranken oder die Schwere der Krankheit. Durch diese „alternierende“ Methode glaubte ich, bei einem

⁵⁴ Dennig 1933 S. 1367.

⁵⁵ Vgl. Arzneimittelgesetz (AMG) § 4 Absatz 23.

⁵⁶ Ob eine juristische Differenzierung des Begriffs interventionelle und nichtinterventionelle Prüfung bereits 1933 im Entstehungszeitraum der vorliegenden Studie angewendet wurde, scheint allerdings fraglich zu sein. In den Reichsrichtlinien zur Forschung am Menschen von 1931 findet sich kein Hinweis auf eine Unterscheidung der Begriffe.

großen Krankengut zu statistisch einwandfreien Resultaten zu kommen (siehe Martini). Gegenüber den Assistenzärzten und Schwestern habe ich die Art des Serums streng geheim gehalten; außer mir wusste niemand darum. Ich halte diesen Umstand für ganz besonders wichtig. Denn nur so wird eine unvoreingenommene Beurteilung gewährleistet.“⁵⁷

Zunächst einmal beschreibt er wie in den Artikeln zuvor ein alternierendes Verfahren, wobei in diesem Beispiel die Zuweisung nicht vom Aufnahmetag, sondern von der Folge des Erscheinens des Patienten in der Klinik abhängig war. Mit seinem direkten Verweis auf Martini begründet er sein Vorgehen damit, dass er ein möglichst einwandfreies statistisches Resultat an einem großen Patientenkollektiv erreichen möchte. Es finden sich in den Literaturangaben keine Hinweise, ob sich Bingel direkt auf die Methodenlehre Martinis bezieht, dennoch ist es nicht unwahrscheinlich.

Das eigentlich Bemerkenswerte an dieser Studie ist allerdings weniger die Beschreibung des alternierenden Verfahrens oder aber der direkte Verweis auf Martini, als vielmehr seine Methode, die Art des Serums, mit welchem die Patienten behandelt wurden, vor dem medizinischen Personal geheim zu halten. Er beschreibt letztendlich nichts anderes als die **Verblindung** eines klinischen Versuchs und begründet sein Vorgehen damit, dass er eine unvoreingenommene Beurteilung der Wirkunterschiede durch die ärztlichen Kollegen und das Pflegepersonal erzielen möchte.⁵⁸ Ob die Patienten von der Art des Serums wussten, beschreibt Bingel nicht. Es ist jedoch unwahrscheinlich, dass die Patienten Kenntnis über die genaue Therapie hatten, das medizinische Personal, das die Medikamente verabreichte, aber nicht. Die differenzierte Stellungnahme von Bingel, warum er sich für eine Verblindung des Versuchs entschlossen hat, macht es wahrscheinlich, dass er auch der Gefahr der voreingenommenen Bewertung durch den Patienten versuchen wollte zuvorzukommen.

Bei der Verblindung eines Versuchsaufbaus wird unterschieden, welche Teilnehmer der Studie „blind“ sind. In dem Text wird explizit erwähnt, dass behandelnde Ärzte und das Pflegepersonal keine Kenntnis von der Art der Medikamente gehabt hätten. Zudem kann man vermuten, dass auch die Patienten „blind“ waren. Da Bingel selbst auch als Untersucher in Erscheinung getreten ist, kann die Versuchsanordnung streng genommen nicht als doppelt verblindeter Versuch betrachtet werden. Auch wenn Bingel sich selbst nicht mit in die „Verblindung“ eingeschlossen hat, kann man den Versuchsaufbau zumindest als eine Art Vorläufer des doppeltblinden Versuchs betrachten.

⁵⁷ Bingel 1949 S. 101.

⁵⁸ Vgl. Bingel 1949 S. 101.

Intention des doppelblinden Versuchs ist es, eine, unvoreingenommene Beurteilung seitens des Patienten und seitens des Arztes zu ermöglichen und potentielle Verzerrungen zugunsten eines Präparates zu vermeiden. Während moderne klinisch-pharmakologische Studien häufig mit diesem Verfahren durchgeführt werden, stellt die Durchführung einer Art Vorläufer des doppelblinden Versuchs 1918 ein sehr innovatives Vorgehen dar. Es finden sich in meinem Material nur wenige andere Studien, in denen eine einfache Verblindung durchgeführt wird, in der nur der Patient keine Kenntnis von der Art des Medikamentes hatte. Doppelblinde Studien wurden nicht gefunden.

Fraglich bleibt, warum Bingel die doppelte Verblindung nicht auch für den 1942 durchgeführten Versuch beschreibt. Im Text finden sich zumindest keine Hinweise, dass er die Versuchsanordnung 1942 ebenfalls verblindet hat. Das ist insofern auffällig, weil er an anderer Stelle die Wichtigkeit der doppelten Verblindung noch einmal explizit herausstellt und andere Arbeiten, in denen auf die Verblindung verzichtet wurde, kritisch hinterfragt: „Ferner hat keiner der Nachprüfer die Art des zu prüfenden Serums vor den Schwestern und den Assistenzärzten geheim gehalten; wenigstens ist dieser Umstand in keiner der erschienenen Arbeiten erwähnt. Und doch muss ich gerade darauf, wie schon betont den allergrößten Wert legen.“⁵⁹

Man kann allein auf der Grundlage der Art, mit der Bingel den Wert der Verblindung beschreibt und wie er andere Kollegen vehement wegen ihrer Methodik kritisiert, nicht automatisch schlussfolgern, dass er die Verblindung auch bei seinen späteren Versuchen 1942 eingesetzt hat.

Auf einen anderen erwähnenswerten Aspekt macht Bingel zum Schluss des Artikels aufmerksam:

„Die letzte Reihe hatte ich zahlenmäßig ausgewertet, als durch Feindeinwirkung eine Anzahl Krankengeschichten vernichtet wurde. Auf Einzelheiten konnte daher nicht mehr eingegangen werden.“⁶⁰

Ein speziell für die Jahre zwischen 1939 und 1945 häufiges Problem sind die Auswirkungen durch den Krieg, die auch für die klinische Forschung schwerwiegende Konsequenzen hatten. Wird in den ersten Kriegsjahren von den Forschern noch relativ selten über kriegsbedingte Hindernisse für ihre Forschung berichtet, so mehren sich speziell in den letzten Jahren des Krieges Hinweise auf immer schwerer werdende Bedingungen für die klinischen Forscher. Auch verschiebt sich das zahlenmäßige Verhältnis von Studien, die in militärischen Einrichtungen durchgeführt wurden,

⁵⁹ Bingel 1949 S. 101.

⁶⁰ Bingel 1949 S. 103.

gegenüber Studien unter zivilen Rahmenbedingungen zugunsten der Studien in militärischen Einrichtungen. Dieser Umstand weist auf eine zunehmende Bedeutung der militärisch-klinischen Forschung hin. Hierfür scheinen im Wesentlichen zwei Gründe verantwortlich zu sein. Zum einen ist die Umsetzung klinischer Forschung abhängig von der Verfügbarkeit entsprechender medizinischer Ressourcen. Durch kriegsbedingten Ressourcenmangel scheint die Durchführung klinischer Forschung deutlich erschwert gewesen zu sein. Zum anderen dürfte auch der klinischen Forschung im militärischen Bereich eine stärkere Unterstützung zugute gekommen sein.

IV. Artikel: Placebokontrolle

Ein weiteres Merkmal moderner klinischer Versuche wendet der Professor für Medizin Ach aus Göttingen an. In dem Versuch, der unter dem Titel „Über die Beeinflussung der geistigen Leistungsfähigkeit durch Recresal“ erschienen ist, soll der Einfluss von Recresal – einem psychomotorisch wirksamen Mittel – auf die geistige Leistungsfähigkeit einiger Probanden untersucht werden. Dabei ging Ach wie folgt vor:

„Die Versuche erstreckten sich über 8 Wochen und zwar wurden abwechselnd 14 Tage Scheinmittel und 14 Tage Recresal gegeben. Diesen Versuchen gingen noch 8 – 12 einübende Versuchstage voraus, so dass jede Versuchsreihe mindestens 9 Wochen in Anspruch nahm. Gearbeitet wurde täglich 40 Minuten, an den Tagen Dienstag bis Freitag 17 bis 21 Uhr, jede Vp (Anmerkung: Abkürzung für Versuchsperson, C. Wolkewitz) stets zur selben Zeit. Die Mittel wurden täglich vormittags und mittags genommen, in den ersten 8 Tagen jeder Periode je 2 Tabletten, später je 3 Tabletten. Auch an den Tagen, an denen kein Versuch stattfand, wurde Recresal bzw. das Scheinmittel gegeben. Der Wechsel von Recresal zum Scheinmittel und umgekehrt fand stets am Samstag statt. Selbstverständlich hatte die Vp keine Kenntnis von dem eigentlichen Zweck der Versuche.“⁶¹

Zuvor erläutert Ach den Inhalt der Versuche:

„Die Aufgabe bestand im Durchstreichen der gleichen Buchstaben im Bourdon-Baadetest, wobei in einer Aufgabe zwei Gruppen von je 6 Buchstaben der Versuchspersonen dargeboten wurden.“⁶²

Der Forscher sollte dann anschließend „...die Fehler und Auslassungen der Probanden registrieren.“⁶³

⁶¹ Ach 1937 S. 290.

⁶² Ach 1937 S. 290.

⁶³ Ach 1937 S. 290.

Achs Verständnis von der methodischen Bedeutung seiner Versuchsanordnung zeigt sich in seiner abschließenden Beurteilung des Versuchs:

„Dass die Recresalwirkung bei unseren Versuchen nicht durch suggestive Einflüsse bedingt sein kann, sei es Autosuggestion oder Fremdsuggestion, ist jedem klar, der einmal eine solche achtwöchige Reihe der zwangsläufigen Arbeitsweise unter Verwendung des Bourdon-Baade-Tests mit dem Wechsel von Scheinmitteln und Recresaltabletten durchgeführt hat. Der Wert unserer Untersuchungen liegt infolgedessen nicht nur in der Feststellung der tiefgehenden Wirkung, welche das Recresal auf die geistige Leistungsfähigkeit ausübt, wir geben vielmehr der Hoffnung Ausdruck, dass die von uns ausgebildete Methode auch zur Untersuchung anderer Heilmittel herangezogen werden möge, die in Bezug auf die seelischen Funktionen eine günstige Wirkung für sich in Anspruch nehmen. Denn gerade für das so wichtige, aber noch wenig geklärte Gebiet der psychischen Wirkung von Arzneimitteln ist die Anwendung exakter, die Leistungsfähigkeit wirklich treffender und die mannigfaltigen Fehlerquellen ausschaltender Methoden besonders vonnöten.“⁶⁴

Sein Appell zur Durchführung *scheinmittelkontrollierter* Studien bezieht sich vor allem auf Forschende aus dem Bereich der Neuro- und Psychowissenschaften. Der Vorteil des scheinmittelkontrollierten Versuchs besteht vor allem in der Erfassung suggestiver Beeinflussung oder Verzerrung des Probanden. Diese Beeinflussung spielt insbesondere in den Neuro- und Psychowissenschaften eine wesentliche Rolle. Ach ist sich dessen offenbar bewusst. Dass aber auch in anderen Disziplinen zum Beispiel in Bereichen der inneren Medizin die suggestive Beeinflussung eine wesentliche Rolle spielt und daher den Einsatz von scheinmittelkontrollierten Versuchen auch in diesen Disziplinen der Medizin sinnvoll macht, diese Schlussfolgerung zieht Ach nicht, sondern beschränkt sich bei seiner Empfehlung für die von ihm durchgeführte Methodik zunächst auf den Bereich der psychisch wirksamen Arzneien. Die Tatsache, dass Ach bereits 1937 eine Art Vorläufer des heute in anderer Form durchgeführten *placebokontrollierten* Versuchs beschreibt, ist ein weiterer Beleg dafür, dass zumindest bei einem Teil der Forscher in Deutschland ein hohes Verständnis für methodische Problemstellungen vorhanden war.

Zwei Besonderheiten des von Ach beschriebenen Versuchs sollen noch erörtert werden. Achs Versuch diente der Erforschung des Einflusses psychotroper Substanzen auf die geistige Leistungsfähigkeit und ist somit dem Bereich der Neurowissenschaften zuzuordnen. Während sich – allgemein gesprochen - in der inneren Medizin Ergebnisse

⁶⁴ Ach 1937 S. 292.

zumeist objektiver darstellen und messen lassen, wie zum Beispiel durch die Messung des Blutdrucks oder durch Bestimmung verschiedener Laborparameter, sind die Ergebnisse insbesondere von Studien, die der Erfassung von psychischen oder mentalen Veränderungen dienen, stärker vom subjektiven Empfinden des Patienten abhängig und verschließen sich häufig einer objektiven Erfassung durch naturwissenschaftliche Parameter. Vor diesem Hintergrund scheint eine fachspezifische methodische Entwicklung bei der Durchführung von Studien unterschiedlicher Disziplinen nicht unwahrscheinlich zu sein.⁶⁵ Insbesondere durch die Einführung von Placebokontrollen und Verblindungen sollen subjektive Verzerrungen seitens des Patienten in Erwartung der Anwendung eines wirksamen Medikamentes erfasst werden. Durch die Kontrolle subjektiver Verzerrungen lässt sich der tatsächliche Einfluss eines Mittels besser gegenüber dem Placeboeffekt abgrenzen. Vor diesem Hintergrund lässt sich auch erklären, warum insbesondere für Studien aus dem Bereich der Psychiatrie oder der Neurologie die Placebokontrolle als auch die Versuchsverblindung sehr sinnvolle Ergänzungen des methodischen Repertoires darstellen. Insbesondere hinsichtlich dieses Aspektes stellt der „Recresal-Versuch“ eine Besonderheit dar.

Zuletzt muss noch darauf hingewiesen werden, dass es sich bei dem Versuch um Untersuchungen an gesunden Probanden handelt, während die anderen in diesem Teil erwähnten Studien an Patienten durchgeführt wurden. Ob diese Tatsache in diesem konkreten Fall tatsächlich einen Einfluss auf die Studiengestaltung gehabt haben könnte, kann retrospektiv kaum geklärt werden. Nur sollte kurz erläutert werden, warum Versuche an gesunden Probanden eine andere Studiengestaltung ermöglichen oder vielleicht sogar erforderlich machen als Versuche an zu behandelnden Patienten. Besteht bei klinischen Versuchen eine Notwendigkeit, die Patienten zu behandeln, fällt diese bei gesunden Personen weg. Zudem muss der Forschende bei einem gesunden Probanden nicht auf den eingeschränkten Gesundheitszustand wie bei einem Patienten Rücksicht nehmen. Diese unterschiedlichen Voraussetzungen geben dem Forschenden unter Umständen mehr Möglichkeiten bei der Studiengestaltung, als bei einer klinischen Studie an Patienten, in der der Behandlungserfolg verfolgt werden muss. Der in dem „Recresal-Versuch“ beschriebene Einsatz von Scheinmitteln ist insgesamt als Ausnahme zu

⁶⁵ Vgl. Kaptchuk 1998 S. 415 :Diese Annahme deckt sich mit der Beobachtung von Kaptchuk. Dieser beschreibt, dass die Einführung der Versuchsverblindung zunächst in Studien aus dem Bereich der Neurowissenschaften erfolgte, bevor die Verblindung auch in pharmakologischen und physiologischen Studien eingesetzt wurde.

betrachten. So wurde bei insgesamt drei von 159 ausgewerteten Studien eine Anwendung von Scheinmittelkontrollen beschrieben.

V. Artikel: Statistische Auswertung

Bezüglich der Studiengestaltung wurden in unterschiedlichen Artikeln verschiedene methodische Grundlagen vorgestellt. Ein weiterer Aspekt, der die Qualität von Studien kennzeichnet, ist die Ergebnisdarstellung und Auswertung. In dem Material variiert die Qualität der Ergebnisauswertung sehr stark. In einigen Studien wird gänzlich auf eine quantitative Darstellung der Ergebnisse verzichtet. In anderen beschränkt sich die Auswertung auf die Angabe von Mittelwerten oder Prozentangaben, die oft als beweisend für die Wirksamkeit eines Präparates dargestellt werden. Die Anwendung komplexer mathematischer Verfahren bei der Ergebnisauswertung findet sich in den untersuchten Studien bei einem insgesamt nur geringen Teil wieder.

In dem Artikel „Vergleichend therapeutische Untersuchungen über die Wirkung von Nerium Oleander auf den gestörten Kreislauf.“ findet sich eines der wenigen Beispiele für eine ausgewogene und detaillierte mathematische Auswertung wieder. In dieser Veröffentlichung werden die Wirksamkeiten von Oleander- und Digitalispräparaten auf den Kreislauf miteinander verglichen. Bei der Ergebnisauswertung definiert der Autor zunächst biologische Parameter, die einen Vergleich zwischen den Präparaten erleichtern sollen:

„Aus den mannigfaltigen Digitalis- bzw. Oleanderwirkungen wählten wir als Vergleichsfaktoren diejenigen aus, die zahlenmäßig klar erfassbar waren. An erster Stelle stehen hier die Ergebnisse der Flüssigkeitsbilanz und die Pulszählung.“⁶⁶

Mit der Auswahl geeigneter Parameter bereitet Freyseng den statistischen Vergleich zwischen den beiden Präparaten vor:

„Bei der statistischen Berechnung folgte ich den Ausführungen Kollers. Durch die Aufstellung des mittleren Fehlers σ erhalten wir ein Maß für die Grenzen des Zufalls. Erst wenn diese bei einem gesicherten Vergleich überschritten sind, darf ein Unterschied als gesichert angesehen werden. (...) σ_x ist also eine Funktion der Differenzen zwischen den Ergebnissen der Einzelbeobachtungen x_1 bis x_n und ihrem arithmetischen Mittelwert x sowie ihrer Anzahl n . Stehen nun 2 Reihen x_1 bis x_n und y_1 bis y_n zum Vergleich, so ergibt sich zunächst ohne weiteres die Berechnung der Mittelwerte x und y , der mittleren Fehler σ_x und σ_y , außerdem die Differenz $(x-y)$. Erforderlich ist jetzt noch der mittlere Fehler der

⁶⁶ Freyseng 1938 S. 816.

Differenz, also $\sigma_{(x-y)}$. (...) Überschreitet $(x - y)$ den dreifachen Wert von $\sigma_{(x-y)}$, so ist ein Unterschied anzuerkennen; bleibt die Differenz innerhalb dieser Grenzen, so ist sie ein Zufallsergebnis.“⁶⁷

In der Publikation lässt sich das Bemühen von Freyseng beobachten, einen statistischen Vergleich zwischen den Präparaten durchführen zu wollen, um Aussagen hinsichtlich von Unterschieden in der Wirksamkeit der Präparate machen zu können. Bei der Auswahl der statistischen Methoden bezieht er sich auf die Vorschläge von Koller und bildet Standardabweichungen, um mögliche Unterschiede zwischen den Präparaten statistisch zu überprüfen. Abschließend kommt er zu dem Ergebnis:

„Ein Unterschied gegenüber Digitalis in der Stärke, dem Eintritt und der Dauer der diuretischen und bradykardischen Wirkung wurde nicht gefunden.“⁶⁸

Die gewissenhafte und detaillierte Darstellung der Anwendung statistischer Methoden zur Ergebnisauswertung in diesem Artikel stellt eine Ausnahme dar. Wie bereits erwähnt, beließen es die meisten Forscher therapeutischer Studien bei einer sehr oberflächlichen Betrachtung der Ergebnisse. Das lässt sich im anschließenden Teil, in dem die quantitativen Ergebnisse der Evaluation vorgestellt werden, verfolgen. Freyseng erklärt und begründet sein Vorgehen bei der Auswertung. Das weist auch darauf hin, dass sich noch keine mathematischen Operationen für die Studienauewertung etabliert haben. Insbesondere ein Meinungsauustausch zwischen Martini und dem oben erwähnten Koller über die Verwendung des mittleren Fehlers, der im Teil „Zur Wahrnehmung von Paul Martini und seiner Methodenlehre“ dargestellt wurde, bekräftigt diese Annahme.

VI. Artikel: Studie aus militärischem Bereich

An dem Beispiel des Artikels „Kombinierte Fleckfieberbehandlung mit Rekonvaleszentenserum und Eigenblut“ soll auf eine Besonderheit speziell der Zeit zwischen 1939 und 1945 eingegangen werden. Der Artikel ist von Raettig veröffentlicht worden, der als Stabsarzt in einem Feldlazarett gearbeitet hat. Raettig schreibt:

„Bei jedem Kranken, der in der Beobachtungsabteilung des Feldlazarettes mit einer fieberhaften Erkrankung aufgenommen wurde und bei dem der berechtigte Verdacht auf Fleckfieber bestand, wurde mit EB.-Gaben (EB – Eigenblut) von je 40 ccm wie im Vorversuch III begonnen. In den wenigen Fällen, in denen sich die Verdachtsdiagnose nicht bestätigte, konnten die EB.-Gaben nicht schaden.“⁶⁹

⁶⁷ Freyseng 1938 S. 816 – 817.

⁶⁸ Freyseng 1938 S. 820.

⁶⁹ Raettig, 1943 S. 562.

Aus dem Text geht hervor, dass die klinischen Versuche in einem militärischen Lazarett durchgeführt worden. Der militärische Charakter einer therapeutischen Studie stellt aus methodischen und ethischen Gründen eine Besonderheit dar. In dem Kapitel „Externe Einflüsse auf die klinische Forschung“ wird detailliert auf das Problem eingegangen. Dennoch sei an solcher Stelle schon einmal auf bestimmte Merkmale dieser Studien hingewiesen.

Die Grundlagen für die Versuchsanordnung unterscheiden sich zunächst einmal bezüglich der Rekrutierung von Versuchspersonen. In militärischen Studien beschränkt sich das Probandenkollektiv auf fast ausschließlich männliche und im Verhältnis zu klinischen Studien junge Patienten. Zudem können äußere Einflussfaktoren wie zum Beispiel Ernährung etc. besser und länger überprüft werden als in zivilen Einrichtungen, in denen erst mit Aufnahme des Patienten eine Erfassung äußerer Faktoren möglich ist. Unter methodischen Gesichtspunkten stellen diese Bedingungen interessante Voraussetzungen für die Durchführung klinischer Studien dar.

Inhaltlich befasst sich der Artikel mit einer therapeutischen Studie zur Behandlung von Infektionskrankheiten. Auch das ist typisch für Studien aus einem militärischen Umfeld zur Kriegszeit.

Speziell bei der Behandlung von Gefangenen, weniger in der Lazarettmedizin, zeichnen sich noch zusätzlich ethische Konflikte der klinischen Forschung in diesem Umfeld ab und stellen unter diesem Aspekt eine Besonderheit unter den evaluierten Studien dar.⁷⁰

Diesen Teil abschließend sollen die wichtigsten Beobachtungen zusammengefasst werden. Aus methodischer Sicht interessant wurde das alternierende Verfahren zur Gruppenbildung kontrollierender Versuche erörtert. Die *zeitliche Alternierung* stellt eine praktikable Möglichkeit dar, Probanden möglichst ohne die Beeinflussung der Neigung des Untersuchers auf verschiedene Studienarme zu verteilen. Das zeitlich alternierende Verfahren kann berechtigterweise als eine Art Vorläufer des heute üblichen randomisierten Verfahrens zur Gruppenbildung verstanden werden.⁷¹ Das Verständnis der Problematik der Verzerrung durch eine von Neigungen gesteuerte Gruppenbildung lässt sich bei einigen Autoren wie zum Beispiel bei Nissen gut verfolgen.

Neben dem alternierenden Verfahren konnte in der Beschreibung des Artikels von Bingel eine Art Vorläufer der *doppelblinden Versuchsanordnung* als methodisches Mittel zur

⁷⁰ Vgl. Roelcke 2007 S. 143.

⁷¹ Vgl. Stoll 2005 S. 106.

Studiengestaltung nachgewiesen werden. Man muss in dem Fall des Blindversuchs von einer Rarität ausgehen, da die Verblindung eines Versuchs in nur fünf Studien explizit Erwähnung findet. Dennoch lässt sich gerade anhand des Artikels von Bingel gut verfolgen, dass das Wissen um die Bedeutung der Verblindung den Lesern der Deutschen Medizinischen Wochenschrift zugänglich gewesen ist. Dass Bingel bereits 1918 die Durchführung eines Vorläufers des modernen doppelblinden Versuchs beschreibt, unterstreicht, dass er unter methodischen Gesichtspunkten seiner Zeit weit voraus war. Der von Ach beschriebene *scheinmittelkontrollierte* Versuch kann als ein weiteres Merkmal des modernen klinischen Versuchs aufgefasst werden. Wie der Blindversuch dient auch die Placebokontrolle der Erfassung und Berücksichtigung der Verzerrung oder, um den englischen Begriff zu verwenden, des „*Bias*“. Placebokontrolle und Verblindung werden in modernen klinischen Studien oft in Kombination eingesetzt. Für die damalige klinisch-therapeutische Forschung stellt die Anwendung der Placebokontrolle eher die Ausnahme dar. So wurden insgesamt in nur drei von 159 Artikeln Placebokontrollen beschrieben.

Die Ergebnisdarstellung und Auswertung klinischer Studien ist ebenfalls kurz angesprochen worden. Werden heute komplexe statistische Methoden bei der Auswertung klinischer Studien eingesetzt, befindet sich die biometrische Analyse von Studien in dieser Zeit noch auf einem sehr niedrigen Niveau. Die Anwendung des *mittleren Fehlers* bei der Beurteilung von Wirksamkeitsunterschieden muss eher als Ausnahme betrachtet werden und ist, wie in dem Kapitel „Zur Wahrnehmung von Martini und seiner Methodenlehre“ deutlich wird, keinesfalls unumstritten.

Die besondere Bedeutung der *Lazarett- und Lagermedizin* für die Gestaltung klinischer Studien wurde erwähnt. Ist der Einfluss der militärischen Umgebung in den Jahren bis 1939 relativ gering, so zeigt sich spätestens ab Kriegsbeginn eine deutliche Zunahme von klinischen Studien aus diesem Kontext. Aus diesem Grund ist eine Auseinandersetzung mit dem Einfluss dieser Arbeiten auf die klinische Forschung notwendig.

3.2. Quantitative Darstellung der Untersuchungsergebnisse

In diesem Kapitel werden verschiedene Ebenen der Studienplanung, -durchführung und -auswertung untersucht und quantitativ dargestellt.

Zunächst soll dargestellt werden, wie viele Studien in den verschiedenen Zeitschriften erschienen sind und wie sie sich auf den Zeitraum zwischen 1933 und 1950 verteilen.

Daran anschließend soll überprüft werden, mit welcher Häufigkeit verschiedene Studientypen auftraten und welche Kontrollmechanismen angewandt wurden.

Aspekte zur Studienvorbereitung, der Durchführung mehrarmiger Studien sowie zur Wahrnehmung von Störgrößen werden in dem darauf folgenden Abschnitt behandelt.

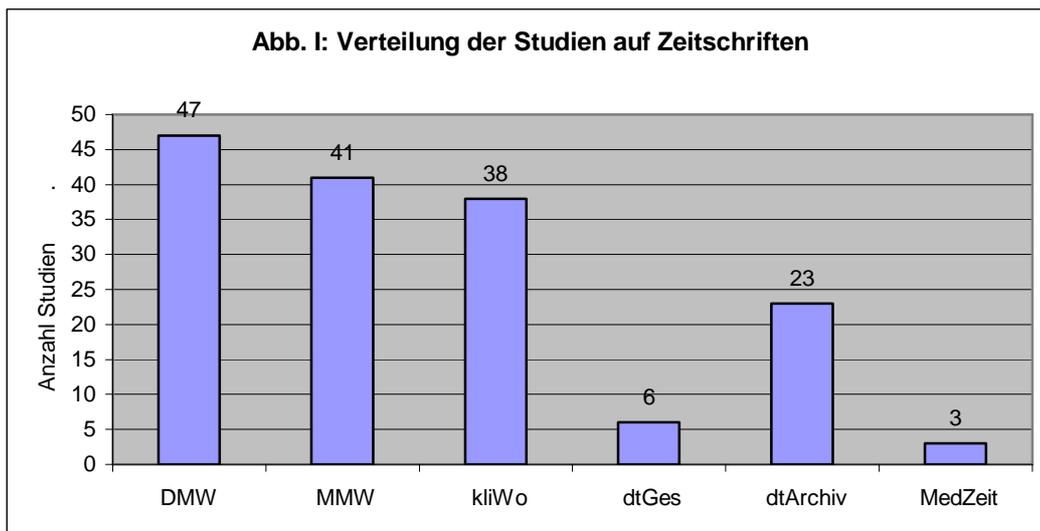
Abschließend werden Merkmale zur Studienauswertung und Diskussion in dem vorliegenden Material untersucht.

1. Studienanzahl und Verteilung

In dem ersten Diagramm wurden die in dem Zeitraum zwischen 1933 und 1950 erschienenen Artikel, die für die genauere Auswertung berücksichtigt wurden, den Zeitschriften zugeordnet, in denen sie veröffentlicht wurde. Man kann vor allem erkennen, dass in den überregional bekannten und auflagestarken deutschen medizinischen Fachzeitschriften, Deutsche Medizinische Wochenschrift, Münchner Medizinische Wochenschrift und der klinischen Wochenschrift sich der Großteil der ausgewerteten Studien befindet. Einen ebenfalls beträchtlichen Beitrag zu dem von mir ausgewerteten Material liefert das deutsche Archiv für klinische Medizin mit insgesamt 23 ausgewerteten Studien. Ein eher kleiner Teil der Studien, insgesamt nur sechs Beiträge, entstammt den Veröffentlichungen aus der deutschen Gesellschaft für Innere Medizin. Drei Studien sind in der Medizinischen Zeitschrift erschienen. In der Medizinischen Zeitschrift wurden die Veröffentlichungen der Deutschen Medizinischen Wochenschrift, der Münchner Medizinischen Wochenschrift, der Klinischen Wochenschrift, der Medizinischen Welt und der Medizinischen Klinik für den Zeitraum zwischen Oktober 1944 und Februar 1945 zusammengefasst.

Folgendes Diagramm zeigt die Verteilung der 158 Studien auf die Zeitschriften:⁷²

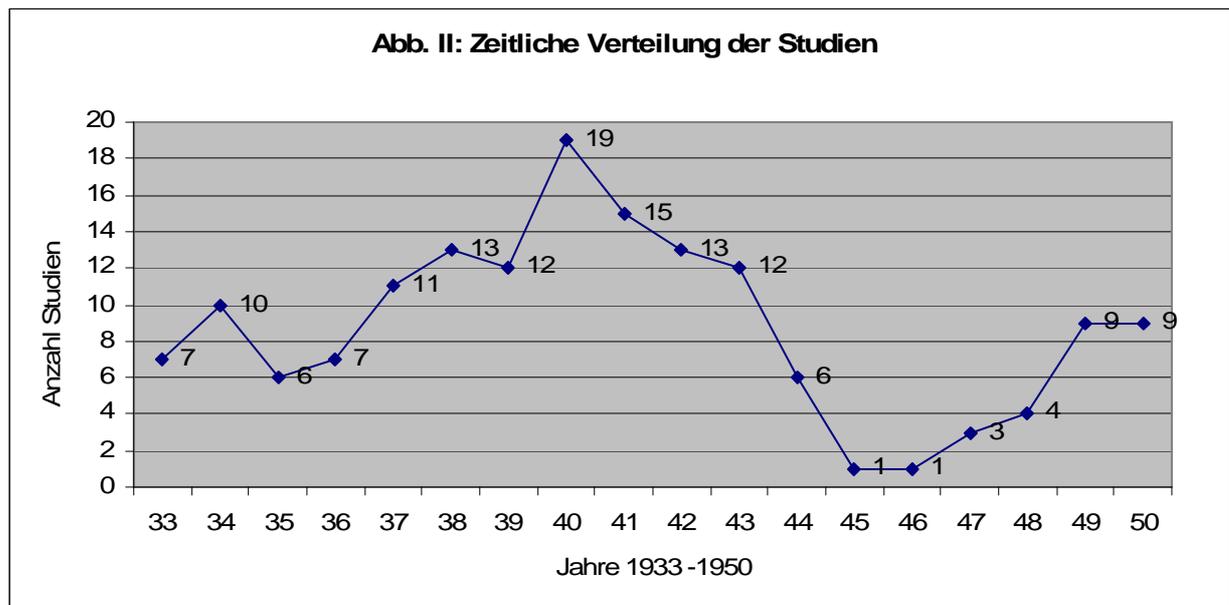
⁷² Erklärung der Abkürzungen der unten dargestellten Abbildung: DMW = Deutsche Medizinische Wochenschrift, MMW = Münchner Medizinische Wochenschrift, kliWo = Klinische Wochenschrift, dtGes = Verhandlungen der Deutschen Gesellschaft für Innere Medizin, dtArchiv = Deutsches Archiv für klinische Medizin, MedZeit = Medizinische Zeitschrift.



Berücksichtigt man, dass in der Deutschen Medizinischen Wochenschrift, der Münchner Medizinischen Wochenschrift und der klinischen Wochenschrift, die aus der Berliner klinischen Wochenschrift hervorgegangen ist, die Gesamtanzahl der Veröffentlichungen pro Jahr um ein Vielfaches höher liegt als in dem Deutschen Archiv für klinische Medizin oder in den Verhandlungen der Deutschen Gesellschaft für innere Medizin, dann zeigt sich hinsichtlich der Konzentration der ausgewerteten Studien in Bezug auf die Gesamtzahl der Veröffentlichungen in den verschiedenen Zeitschriften ein etwas anderes Bild.

Insbesondere in der Zeitschrift Deutsches Archiv für klinische Medizin findet sich bezogen auf die Anzahl der Gesamtveröffentlichungen ein überproportional hoher Anteil evaluierter Studien. Diese Tatsache kann durchaus als Hinweis darauf verstanden werden, dass in der Zeitschrift „Deutsches Archiv für klinische Medizin“ höhere qualitative Ansprüche an Veröffentlichungen gestellt wurden. Denn, wie bereits erläutert, war ein wichtiges Kriterium, das die Auswahl der Studien leitete, die Qualität der Veröffentlichung. Dabei war vor allem wichtig, wie präzise die Versuchsanordnung beschrieben wurde und ob eine wissenschaftliche Fragestellung bei der Planung der Versuchsanordnung formuliert wurde. Die Beobachtung, dass die Zeitschrift Deutsches Archiv für klinische Forschung eine unter wissenschaftlichen Ansprüchen herausgehobene Rolle unter den anderen Zeitschriften hatte, deckt sich mit Eindrücken, die bei der groben Durchsicht der Zeitschriftenbände gewonnen wurde.

Ordnet man die ausgewerteten Artikel dem Erscheinungsjahr zu, ergibt sich folgendes Diagramm:



Anhand des Diagramms lässt sich eine bemerkenswerte Entwicklung verfolgen. Es ist zu beobachten, wie die Anzahl ausgewerteter Veröffentlichungen von einem mäßigen Niveau in der Vorkriegszeit einen Höhepunkt 1940 erreicht. In den letzten Kriegsjahren zwischen 1943 und 1944 fällt die Zahl der ausgewerteten Studien rapide, um in den Jahren 1945 und 1946 mit je einer therapeutischen Studie einen Tiefpunkt zu erreichen. Nach 1945 braucht es vier Jahre bis 1949 wieder das Vorkriegsniveau von 1938 erreicht werden kann.

Der massive Einbruch der Anzahl ausgewerteter Arbeiten ist mit Sicherheit als Folge des Krieges zu betrachten. Einige Blätter mussten nach 1944 die Veröffentlichung ganz aufgeben. So musste der Druck der Zeitschriften Deutsche Medizinische Wochenschrift, die Klinische Wochenschrift und Münchner Medizinische Wochenschrift unterbrochen werden. An ihre Stelle trat für die Zeit zwischen Oktober 1944 und Februar 1945 die Medizinische Zeitschrift. Der Druck der beiden kleineren Blätter Deutsches Archiv für klinische Medizin und die Zeitschrift „Verhandlungen der Gesellschaft für Innere Medizin“ musste für das Jahr 1945 ersatzlos eingestellt werden, konnte ein bis zwei Jahre später aber wieder aufgenommen werden. Die Münchner Medizinische Wochenschrift konnte erst wieder ab 1950 verlegt werden. Erklärlich wird die Einstellung des Drucks der Zeitschriften vor dem Hintergrund des Krieges, der spätestens ab 1943 durch die zunehmende Zerstörung der Infrastruktur in Deutschland zu einem massiven und allgemeinen Ressourcenmangel geführt hat.

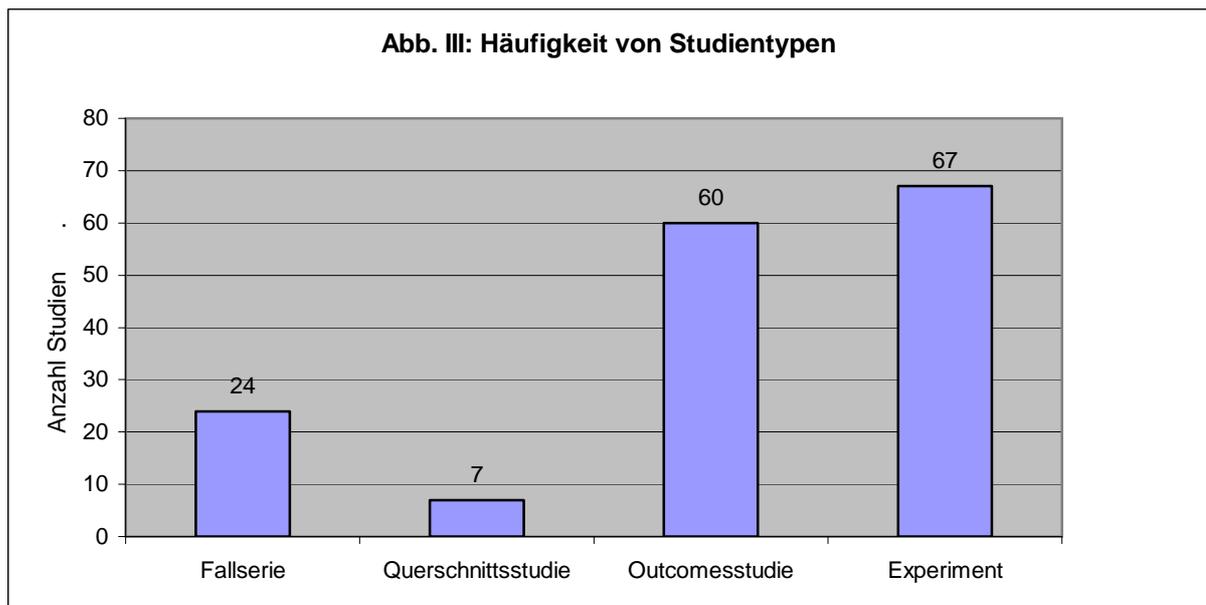
Während das Nichterscheinen medizinischer Zeitschriften und somit auch klinischer Studien zum Ende des Krieges gut zu erklären ist, ist es weniger einleuchtend, dass 1940 ein Höhepunkt der Anzahl ausgewerteter Studien zu verzeichnen ist und dass es nach 1945

relativ lange braucht, bis das Vorkriegsniveau wieder erreicht wird. Sicherlich wird schon allein das Nichterscheinen der „Münchner Medizinischen Wochenschrift“ zwischen 1945 und 1949 eine deutliche Abnahme der ausgewerteten Studien bewirkt haben. Dennoch ist es fraglich, ob sich die Abnahme der Anzahl ausgewerteter Studie nach 1945 dadurch schon ausreichend erklären lässt. Bestimmt werden auch die Zerstörung wichtiger klinischer und journalistischer Einrichtungen und der allgemeine Ressourcenmangel nach 1945 einen Beitrag zur Erklärung liefern können. Ein zumindest aber ebenso diskussionswürdiger Ansatz zur Erklärung dieser Beobachtung könnte auch darin bestehen, dass sich das Interesse an klinischer Forschung verändert hat, oder aber dass sich die ethischen Grundlagen klinisch-therapeutischer Forschung in Deutschland verschoben haben.⁷³

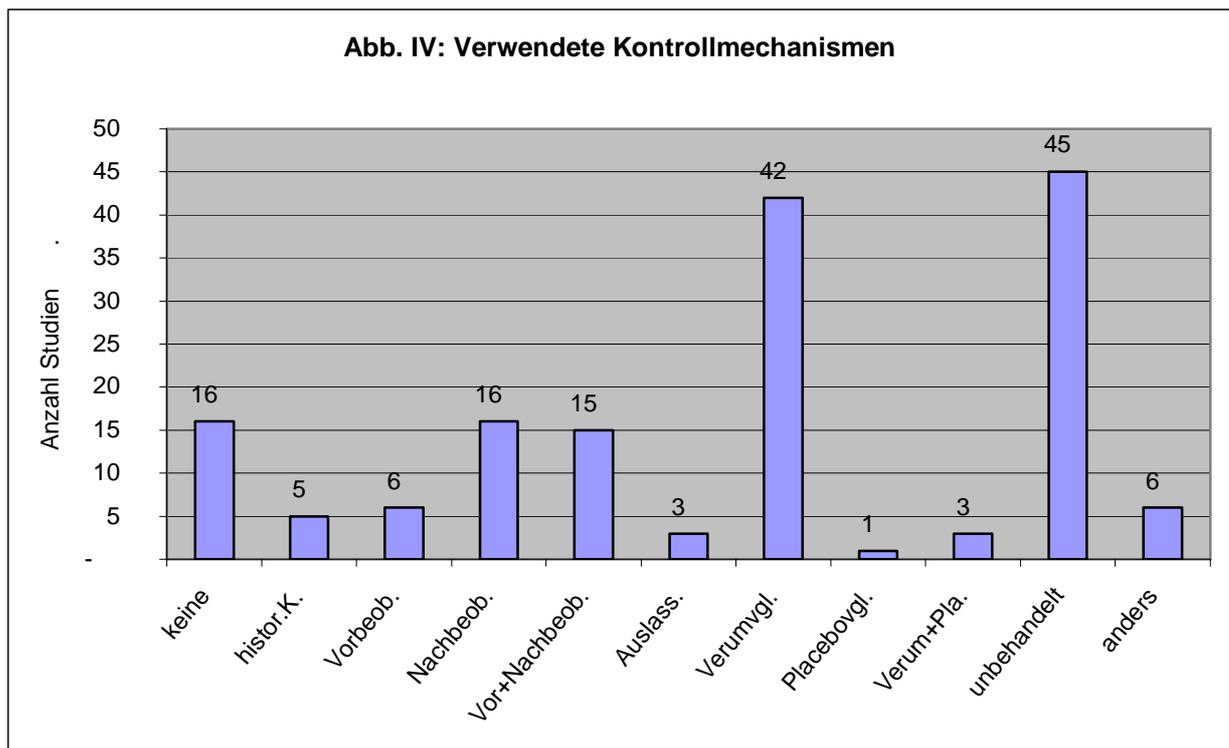
2. Studientypen und Kontrollmechanismen

Im nächsten Diagramm wurden die Studien hinsichtlich ihrer angewandten Methodik vier verschiedenen Kategorien zugeordnet. Bei der Zuweisung der Studien zu den Kategorien möchte ich auf die in dem Kapitel „Material und Methoden“ erläuterte Herangehensweise verweisen. Anzumerken ist noch, dass auch in diesem Fall eine eindeutige Zuweisung der Studien zu einer der Kategorien aufgrund zum Teil lückenhafter Darstellungen nicht immer leicht fiel. Aus der Darstellung geht hervor, dass in 67 von 158 ausgewerteten Studien eine experimentelle Versuchsanordnung in Form von Kontrollgruppen gebildet wurde. Mit umgerechnet 42 % machen Studien mit einem experimentellen Charakter den größten Anteil der vier gebildeten Kategorien aus. Die zweitgrößte Gruppe stellt die Kategorie der Outcomesstudien mit einer Gesamtzahl von 60 oder einem Anteil von 38 % dar. Die Gruppe der Fallserie macht einen Anteil von 15 % aus. Mit fünf % am seltensten konnte eine Querschnittsstudie aufgefunden werden.

⁷³ Hinweise auf veränderte Rahmenbedingungen klinischer Forschungen nach dem zweiten Weltkrieg finden sich unter anderem auch bei Tröhler (1993 S. 73) und Deutsch (1997 S. 108).



Bei der Betrachtung und Auswertung der Ergebnisse dieser Aufstellung muss auf einen wichtigen Aspekt hingewiesen werden. Es wurde bereits erläutert, dass die Definition des Begriffs „klinische Studie“ und die Abgrenzung von Studie zu unsystematischen Erfahrungsbeobachtungen mit Schwierigkeiten behaftet sind. Als wesentliches Abgrenzungskriterium wurde in dieser Arbeit die Frage nach einer „ex ante“ – Planung, also einer Vorausplanung definiert. Bei der Auswahl der Arbeiten wurden also Veröffentlichungen bevorzugt berücksichtigt, bei denen eine systematische Versuchsplanung ersichtlich ist. Während eine Versuchsplanung mit einem experimentellen Charakter, also mit zwei verschiedenen behandelten Gruppen, ohne eine systematische Planung kaum denkbar ist und daher eine Verwechslung mit unsystematischen, ungeplanten Erfahrungsberichten weniger möglich ist, ist die Verwechslungsgefahr mit Outcomesstudien oder Fallserien tendenziell wahrscheinlicher. Durch die besondere Berücksichtigung des Merkmals „Vorausplanung“ ergibt sich möglicherweise eine leichte Verzerrung zugunsten der Studienform des Experimentes. Aufgrund dieser potentiellen Verzerrung wurden sämtliche Studien zweimal beurteilt. Insgesamt lässt sich aus der Darstellung verfolgen, dass in der klinischen Forschung die Anwendung von Experimenten ein gebräuchliches Vorgehen darstellte. Um zu zeigen, welche Kontrollmechanismen in welchem Maß angewandt wurden, wurde die absolute Häufigkeit den beschriebenen Kontrollmechanismen in folgendem Diagramm zugeordnet:



Aus dem Diagramm geht hervor, dass in 16 der 158 evaluierten Studien kein Kontrollmechanismus beschrieben wurde. Dies entspricht einem Anteil an dem ausgewerteten Material von ungefähr 10 %. Die Kontrollmechanismen, die von den Autoren am häufigsten beschrieben wurden, sind mit einem Anteil von 28 % die unbehandelte Kontrollgruppe und mit 26 % die Verumkontrolle. Die Kontrollmechanismen der Vor- oder Nachbeobachtung und der kombinierten Vor- und Nachbeobachtung werden in insgesamt 37 Arbeiten erwähnt, was einem Anteil von ungefähr 24 % entspricht. In nur fünf Studien (3 %) werden historische Kontrollen erwähnt. Sehr selten – in nur vier Arbeiten oder in 3 % aller Fälle - konnte in den Studien eine Placebo- oder eine kombinierte Verum- Placebokontrolle nachgewiesen werden.⁷⁴

Ein weiteres evaluiertes Merkmal der Studien war die Frage nach dem Zeitraum, in dem die Studien durchgeführt worden. In 60 (38%) Arbeiten fanden sich keine Hinweise, bezüglich der Zeitspanne, in der die Studiendurchführung erfolgte. Bei den Studien, in denen sich direkte oder indirekte Hinweise auf die Studiendauer fanden, kann bei einem großen Teil – 54 Studien oder 34 % - davon ausgegangen werden, dass für die Durchführung der Studie eine Zeitspanne zwischen einem und elf Monaten benötigt wurde. In 35 (22 %) Veröffentlichungen wurde ein Zeitrahmen von mehr als einem Jahr

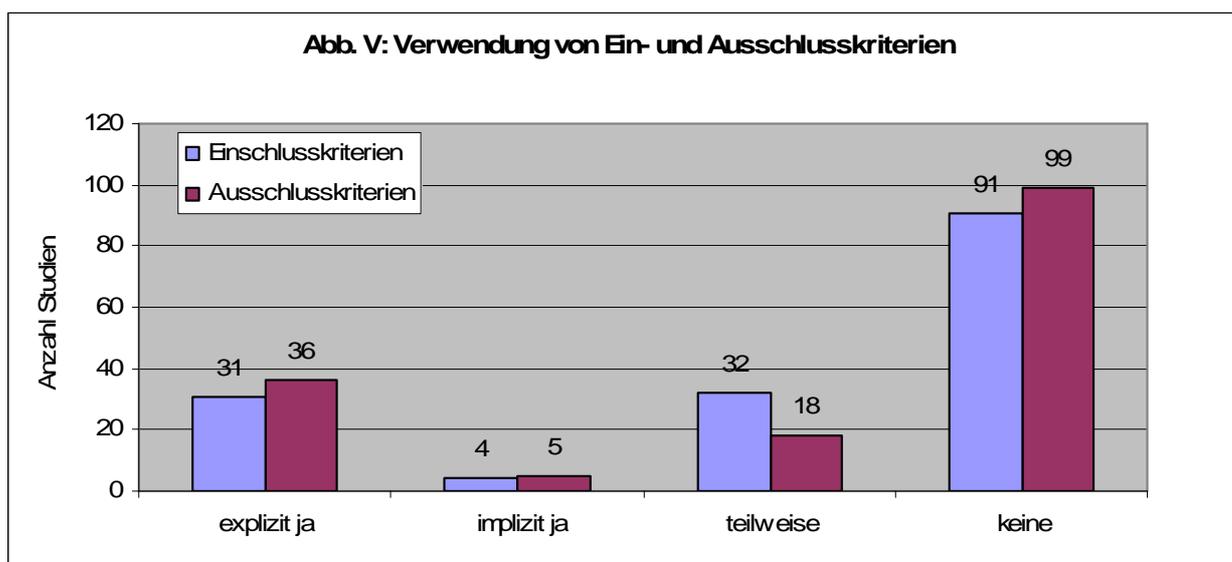
⁷⁴Anmerkung: Für Studien, in denen mehrere Kontrollmechanismen gleichzeitig angewandt wurden, wurde nur der im Vordergrund des Versuchs stehende Mechanismus gewertet, so dass Mehrfachwertungen pro Studie ausgeschlossen sind.

festgestellt. Es fanden sich ferner eine Studie mit einer Dauer von einem Tag, fünf Studien mit einer Dauer zwischen einer und vier Wochen und drei Studien mit einer Dauer länger als fünf Jahre.

3. Studienplanung und Durchführung

Im Folgenden möchte ich auf die detaillierte Untersuchung der Studiengestaltung eingehen. Des besseren Verständnisses und der Übersichtlichkeit wegen bietet es sich an, die Darstellung in verschiedene Themenkomplexe zu gliedern. Dabei wird auf die in dem Kapitel „Material und Methoden“ erwähnte Gliederung zurückgegriffen. Zunächst wird auf das Thema der **Studienvorbereitung und Planung** eingegangen. Im zweiten Teil werden die Ergebnisse erläutert, die vor allem den Bereich der **mehrmarmigen Studiengestaltung** betreffen. Darauf folgend wird auf die Wahrnehmung von **Störgrößen** Bezug genommen. Viertens schließen sich die Fragen nach der **kritischen Auseinandersetzung, der Art der Auswertung und der Wahrnehmung ethischer Probleme** an.

Bei dem ersten Punkt ist das interessierende Merkmal die **Studienvorbereitung**. Hierbei geht es vor allem um die Frage nach der Rekrutierung der Patienten. Dieses Merkmal lässt sich gut abbilden, indem man die Arbeiten hinsichtlich der Beschreibung von Ein- oder Ausschlusskriterien und eines Stichprobenplans untersucht. Dementsprechend ergab sich für die Frage nach der Formulierung von Ein- oder Ausschlusskriterien folgende Verteilung:

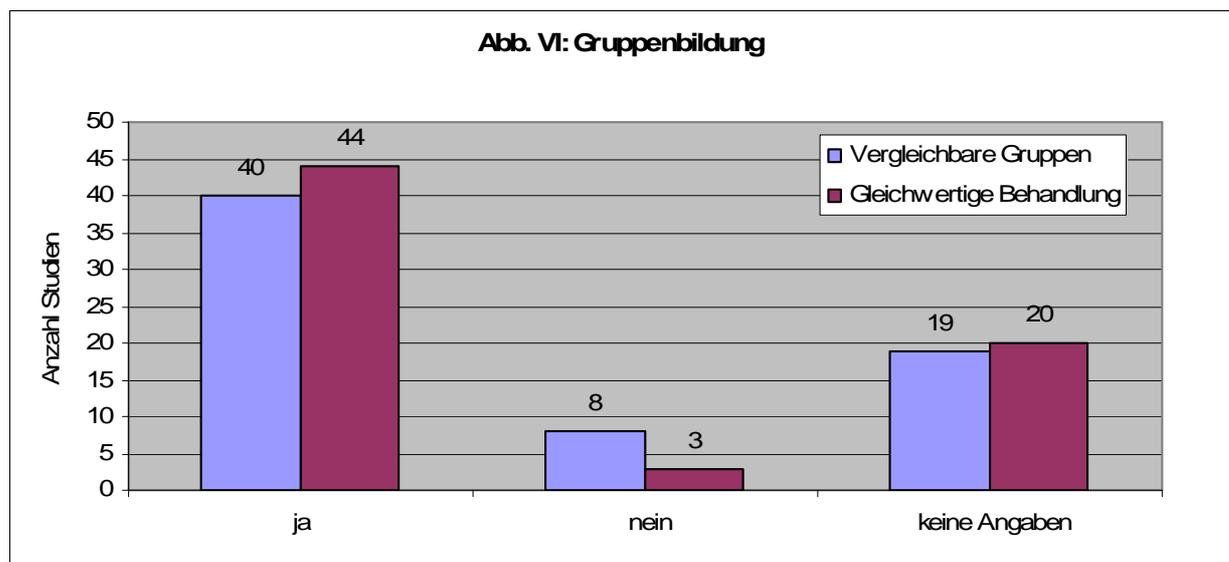


Man kann erkennen, dass in 31 Studien die explizite Beschreibung von Einschlusskriterien formuliert wurde. Bei vier Studien lässt sich durch die Darstellung der Studiendurchführung implizit ableiten, dass Kriterien, die zum Einschluss von Probanden

führten, vorhanden waren. In 32 weiteren Studien kann man davon ausgehen, dass teilweise Einschlusskriterien bei der Patientenrekrutierung behilflich waren. Der Begriff „Teilweise“ bezieht sich darauf, dass in einigen Studien auf die Verwendung von Kriterien zurückgegriffen wurde, diese aber entweder sehr allgemein formuliert waren, wie zum Beispiel „*männliche Versuchsprobanden*“, oder aber sich aus der Erwähnung von Ausschlusskriterien ableiten ließen, wie zum Beispiel durch die Angabe von Altersbeschränkungen. Bei der überwiegenden Mehrzahl der untersuchten Veröffentlichungen finden sich allerdings keine Anhaltspunkte, dass bei der Patientenrekrutierung Einschlusskriterien angewandt wurden. Bei der Darstellung der Verwendung von Ausschlusskriterien präsentiert sich ein ähnliches Bild. Auch hier zeigt sich, dass in der überwiegenden Mehrzahl der Veröffentlichungen auf die Erwähnung von Ausschlusskriterien verzichtet wurden. In 17 Studien wurden sowohl Ein- und Ausschlusskriterien explizit erwähnt. Bei der Frage, ob von den Autoren ein Stichprobenplan beschrieben wurde, konnte in 14 Studien eine explizite Erwähnung nachgewiesen werden. In 19 Studien kann man implizit von der Bildung eines Stichprobenplans ausgehen, während sich bei den 125 übrigen Studien keine Beschreibung oder keine eindeutigen Hinweise auf die Verwendung eines Stichprobenplans finden ließen. Natürlich muss darauf verwiesen werden, dass das Feststellen selbst, dass in der überwiegenden Anzahl der Veröffentlichungen weder Ein- oder Ausschlusskriterien noch die Bildung eines Stichprobenplans erwähnt wurde, kein hinreichender Beleg für das tatsächliche Nichtvorhandensein dieser Kriterien ist. In einem großen Teil der Studien, in denen keine Kriterien definiert wurden, darf man sicherlich annehmen, dass aufgrund einer mangelnden Planung tatsächlich keine Ein- oder Ausschlusskriterien oder ein Stichprobenplan formuliert wurden. In anderen Fällen muss man aber auch in Betracht ziehen, dass Ein- und Ausschlusskriterien und Stichprobenpläne im Text nicht beschrieben wurden, aber dennoch existierten und angewendet wurden. Eine kritische und differenzierte Darstellung der Studienplanung scheint für die meisten zeitgenössischen Autoren dieser Arbeiten keinesfalls als selbstverständlich zu gelten. Vielmehr zeigt sich, dass nur ein eher kleiner Teil der Autoren bewusst auf die Studienplanung eingegangen ist.

Im Folgenden werden verschiedene Aspekte der Bildung von Kontrollgruppen untersucht. Dabei interessiert vor allem, auf welche Art die Gruppenbildung zustande gekommen ist, ob die Autoren bei der Bildung auf *Vergleichbarkeit* der Gruppen geachtet haben und inwieweit eine ansonsten studienunabhängige Gleichbehandlung der Gruppen erwähnt

wurde. Das Merkmal der Vergleichbarkeit der Gruppen lässt sich am besten an den Kriterien Anzahl der Teilnehmer, Geschlechter- und Altersstruktur in den Gruppen überprüfen. Geht zum Beispiel aus der Veröffentlichung hervor, dass die Anzahl der Teilnehmer und die Geschlechter- oder Altersstruktur in den Gruppen etwa gleich war, so kann man wenigstens von einer partiellen Vergleichbarkeit zwischen den Gruppen ausgehen. Von den 67 Studien mit einem experimentellen Versuchsaufbau ergab sich hinsichtlich der Merkmale der Vergleichbarkeit zwischen den Gruppen und der Gleichbehandlung der Gruppen folgende Verteilung:



In 40 von 67 Fällen wurde von den Studienautoren beschrieben, dass die Gruppen vergleichbar waren. In acht Studien wurde behauptet, die Gruppen wären nicht vergleichbar oder aber es ging aus dem Text hervor, dass diese nicht vergleichbar sein konnten. In 19 Studien konnten keine verwertbaren Aussagen oder Hinweise gefunden werden, die auf eine Vergleichbarkeit zwischen den Gruppen hindeuten.

Die Studienteilnehmer wurden überwiegend gleichwertig behandelt. Nur in drei Studien ergaben sich eindeutige Belege dafür, dass die Gruppen nicht gleichwertig behandelt wurden. In 20 Studien konnte diese Frage nicht hinreichend geklärt werden. Für die Studien, in denen keine Kontrollgruppe gebildet wurde, fanden sich in 39 von 92 Studien Hinweise darauf, dass die Studienteilnehmer eine gleichwertige Behandlung bekamen. In einer ebenso großen Anzahl konnte keine eindeutige Aussage diesbezüglich gemacht werden und aus 17 Veröffentlichungen geht hervor, dass die Behandlung als nicht gleichwertig betrachtet werden kann.

Im Bereich der kontrollierten Studien wurde außerdem untersucht, ob und welche **Kriterien der Gruppenbildung** zugrunde gelegt worden sind. Unter den Arbeiten, bei

denen Kriterien zur Gruppenbildung beschrieben wurden, wurde weiter untersucht, ob ein alternierendes Verfahren bei der Gruppenbildung eingesetzt wurde. Außerdem wurde in allen Studien untersucht, ob eine Verblindung der Behandlung durchgeführt wurde.

In 35 (52 %) von 67 Studien ließ sich nicht verfolgen, auf welcher Grundlage eine Gruppenbildung zustande gekommen war. In 32 Studien (48 %) wurden Kriterien formuliert oder es fanden sich eindeutige Hinweise für die Anwendung von Kriterien. In 23 von insgesamt 67 kontrollierten Studien (34 %) wurde ein alternierendes Verfahren bei der Gruppenbildung beschrieben. Das alternierende Verfahren wurde von Stoll als Quasi-Randomisierung⁷⁵ bezeichnet und stellt eine Art Vorläuferverfahren der heute verwendeten Randomisierung dar.

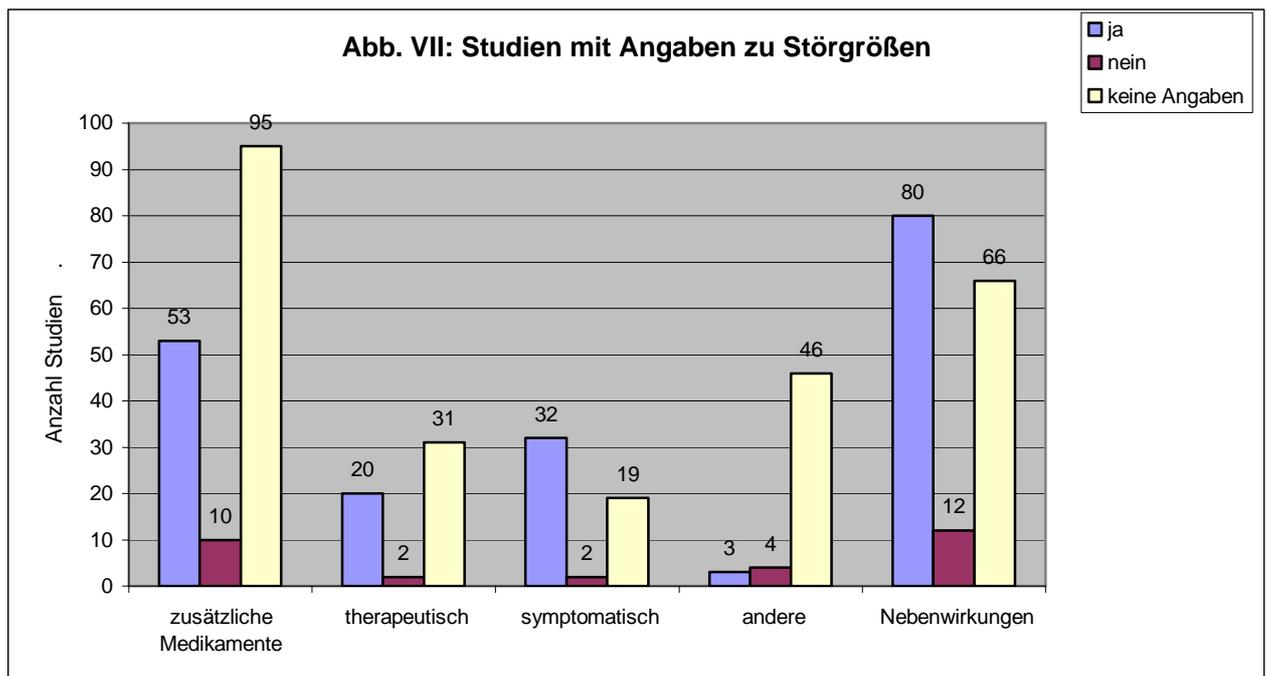
In acht (5 %) von 158 Studien wurde eine Verblindung der Behandlung beschrieben. Es fand sich keine Studie, in der eine doppelte Verblindung der Versuchsanordnung beschrieben wurde. Lediglich in einer Veröffentlichung von Adolf Bingel findet sich eine Erwähnung eines bereits 1918 durchgeführten Versuchs, in dem eine Art Vorläufer einer doppelblinden Versuchsanordnung angewandt worden ist.⁷⁶

Ein weiterer Aspekt, der untersucht wurde, befasst sich mit der Beschreibung von Störgrößen in den Arbeiten. Für die Frage, ob *Störgrößen* in den Studien vorhanden waren, ist es wichtig entsprechende Merkmale zu definieren. Als ein passendes Merkmal dafür kann die Erwähnung von zusätzlichen Medikamentengaben betrachtet werden, da sich dieser Punkt in den Veröffentlichungen relativ gut verfolgen lässt. Unterschieden wurde zwischen Medikamenten, die symptomatisch oder kurativ eingesetzt wurden. Medikamente, bei denen nicht sicher festgestellt werden konnte, welchem Zweck sie dienten, wurden in der Kategorie „andere“ eingeordnet. Außerdem wurde evaluiert, in welchem Maß Nebenwirkungen, der zu untersuchenden Therapie wahrgenommen wurden. Die Erfassung von Nebenwirkungen der zu prüfenden Medikamente ist in diesem Zusammenhang unter zwei Gesichtspunkten zu beachten. In den meisten Studien diente die Beschreibung von Nebenwirkungen dem Zweck, mögliche Einschränkungen und Kontraindikationen beim Einsatz des zu prüfenden Mittels zu erfassen. Zum anderen wurde aber auch der negative Effekt von Nebenwirkungen auf die Durchführung von Studien diskutiert. Unter diesem Aspekt wurden Nebenwirkungen auch als Störgrößen wahrgenommen.

Im folgenden Diagramm sind die Ergebnisse zusammengefasst:

⁷⁵ Vgl Stoll 2005 S. 106.

⁷⁶ Bingel 1949 S. 101.



In 53 (33 %) Studien wurde beschrieben, dass eine zusätzliche Medikation erfolgte. Davon wurde in 45 Arbeiten explizit eine zusätzliche Medikamentengabe erwähnt, in acht Studien implizit. In zehn (sechs %) Arbeiten wurde eine zusätzliche Medikamentengabe verneint, von denen wurde in acht Fällen explizit darauf hingewiesen, in zwei Fällen konnte implizit auf ein solches Vorgehen geschlossen werden. In 94 (61 %) Veröffentlichungen fanden sich keine Hinweise auf die Verwendung oder Nichtverwendung zusätzlicher Medikamente. Unter den 53 Arbeiten, bei denen eine zusätzliche Medikation ausgewiesen wurde, wurde in 20 Fällen ein kurativer Nutzen verfolgt, während sich in 31 Fällen keine Hinweise darauf ergaben. In 32 Studien fanden sich Angaben, dass die zusätzlich verabreichten Medikamente zur symptomatischen Therapie eingesetzt wurden. In 19 Fällen fanden sich dazu keine Angaben. Additiv wurden vor allem Schmerzmittel verabreicht. In drei Studien wurden Medikamente aus anderen zum Teil nicht näher erläuterten Gründen gegeben. In insgesamt neun (6 %) Studien wurden Medikamente sowohl aus kurativen als auch aus symptomatischen Gründen gegeben.

Nebenwirkungen der Therapie wurden in 80 (51 %) der Studien diskutiert. In zwölf (8 %) Arbeiten wurde das Auftreten von Nebenwirkungen verneint und in 66 (41 %) Veröffentlichungen fanden sich keine Angaben diesbezüglich.

Insbesondere bei der Erfassung von Nebenwirkungen konnte bei einem Großteil der Studien festgestellt werden, dass diese erörtert wurden. Dabei reicht die Art der Diskussion von knapp formulierten Feststellungen von Nebenwirkungen bis hin zu detailliert

beschriebenen Nebenwirkungen, deren Konsequenzen für die Durchführung der Studien kritisch erläutert wurden.

Bei den zehn Studien, in denen eine zusätzliche Medikamentengabe grundsätzlich ausgeschlossen werden konnte, ließ sich verfolgen, dass offenkundig durch den bewussten Verzicht von zusätzlichen Medikamenten ein Verständnis für die verzerrende Wirkung zusätzlicher Medikamentengabe vorhanden war. Dass in einigen Studien eine studienunabhängige Medikation weiter durchgeführt wurde, kann allerdings nicht zwingend als Hinweis auf das Unverständnis der potentiellen Ergebnisverzerrung durch andere Störgrößen gewertet werden, da durch bestimmte klinische Situationen der Einsatz von zusätzlichen Medikamenten erforderlich wird. Vielmehr ist wichtig, ob sich die Autoren dieser Störvariablen bewusst waren. So kann man eher davon ausgehen, dass in den Studien, in denen die Verwendung von zusätzlichen Medikamenten beschrieben wurde, eine Darstellung der Störfaktoren beabsichtigt wurde und in den Arbeiten, in denen sich keine Angaben dazu fanden, dieses Problem nicht erkannt beziehungsweise vernachlässigt wurde.

4. Studiauswertung und Darstellung

In dem folgenden Teil soll auf zwei wesentliche Punkte eingegangen werden. Zum einen ist es wichtig zu überprüfen, ob die *Ergebnisdarstellung* adäquat erfolgt ist. Andererseits soll der Frage nachgegangen werden, inwiefern die Autoren ihre Ergebnisse hinsichtlich *methodischer* und *ethischer Fragestellungen* hinterfragt haben.

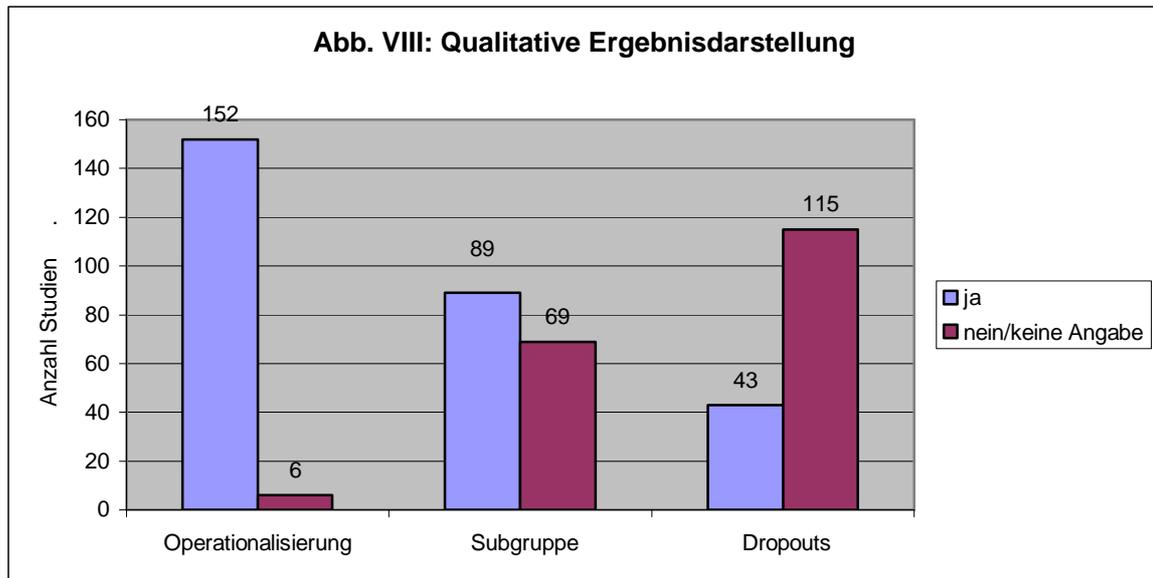
Um zu beurteilen, ob eine adäquate Ergebnisdarstellung erfolgt ist, wurden verschiedene Parameter definiert, die anhand des vorhandenen Materials gut zu evaluieren sind. Bevor Ergebnisse gemessen werden können, sollten Ergebnisparameter operational definiert werden. Die Operationalisierung beschreibt, ob interessierende Merkmale eindeutig definiert wurden und wie sie valide und reliabel gemessen werden sollen.⁷⁷ Weiter stellt sich die Frage, ob eine sinnvolle statistische Darstellung der Ergebnisse durchgeführt wurde und wie komplex die statistische Auswertung erfolgt ist. Eine Subgruppenanalyse⁷⁸ oder die Erwähnung von Probanden, die aufgrund verschiedener Ursachen von der

⁷⁷ Anmerkung: Um zum Beispiel die Wirkung von Antihypertensiva zu prüfen, ist die systematische Erfassung und der Vergleich von Blutdruckwerten als Operationalisierung zu verstehen.

⁷⁸ Vgl. Fletcher 1999 S. 215.

Ergebnisauswertung ausgeschlossen wurden⁷⁹ - kurz „Dropouts“ - kann ebenfalls als ein Gütemaß der Ergebnisauswertung betrachtet werden.

Stellt man die Ergebnisse bei der Evaluation der Artikel hinsichtlich der Punkte **Operationalisierung**, **Subgruppenbildung** und **Dropouts** dar, zeigt sich folgendes Diagramm:



In 152 (96 %) von 158 Studien wurden die Ergebnisse operationalisiert. Nur in sechs Studien konnte bei der Evaluation keine Operationalisierung erkannt werden. Eine Subgruppenbildung wurde in 89 (56 %) von 158 Studien beschrieben. In zwei Studien ergaben sich Hinweise darauf, dass die Patienten nicht ihrer Gruppe zugeordnet wurden. Das Bemühen, homogene Versuchsgruppen zu bilden, spiegelt sich auch in vielen diesbezüglichen Anmerkungen zu methodischen Fragestellungen seitens der Autoren wieder und man kann davon ausgehen, dass das Verständnis der Bedeutung der Subgruppenbildung tendenziell bei vielen zeitgenössischen Autoren vorhanden war. In 37 von 43 Studien wurden explizit Dropouts erwähnt und in sechs Studien konnte aus der Versuchsbeschreibung auf das Vorliegen von Dropouts geschlossen werden. Auch in diesem Fall lässt sich nicht zweifelsohne klären, ob in den 115 Studien, die weder explizit noch implizit Dropouts beschreiben, nicht doch auch Patienten oder Probanden verloren oder von der weiteren Ergebnisauswertung ausgeschlossen wurden. Im Vergleich zu dem Kriterium der Subgruppenbildung scheint es allerdings ein weniger angewandtes Instrument zur Homogenisierung der Gruppen zu sein.

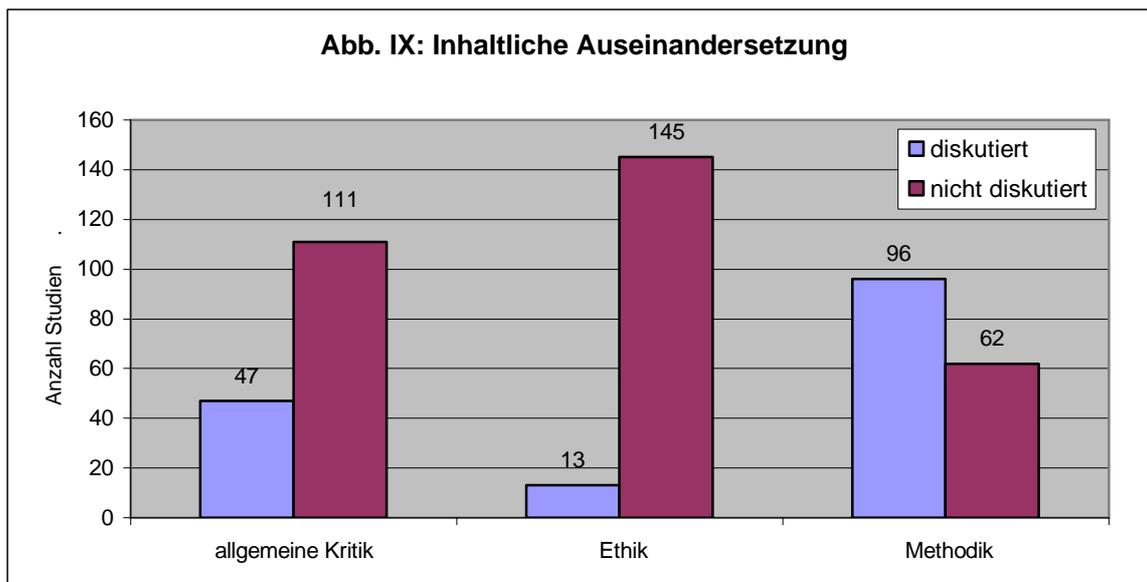
Die Art der **statistischen Auswertung** einer Studie kann ebenfalls als ein Gütekriterium verstanden werden. In einem in der Deutschen Medizinischen Wochenschrift

⁷⁹ Vgl. Fletcher 1999 S. 203.

veröffentlichten Meinungsaustausch zwischen Martini, Heidenhain sowie Koller geht es um die Verwendung und Berechnung statistischer Werte zur Beurteilung klinischer Resultate. Ich werde an anderer Stelle auf diesen Meinungsaustausch noch näher eingehen.⁸⁰ In der modernen klinischen Forschung hat die genaue statistische Darstellung komplexer medizinischer Fragen eine besondere Bedeutung. In den ausgewerteten Studien finden sich durchaus einige Beispiele, die auf ein differenziertes Verständnis statistischer Probleme hinweisen, und auch der oben erwähnte Meinungsaustausch bestätigt diesen Eindruck. Doch in den meisten Studien zeigt sich ein anderes Bild, in dem die Autoren zumeist eine einfache deskriptive statistische Ergebnisauswertung vor komplexeren statistischen Verfahren bevorzugt haben. So wurde in 93 (59 %) von 158 Studien die Anwendung einfacher deskriptiver Auswertungen beschrieben. Unter einer einfachen deskriptiven Auswertung wurde zum Beispiel die Berechnung von Mittelwerten verstanden. Die Berechnung komplexerer, statistischer Werte zur Ergebnisauswertung, wie zum Beispiel die Berechnung des mittleren statistischen Fehlers, wurde in 19 (12 %) der ausgewerteten Arbeiten dargestellt. In 46 (29 %) Studien wurden entweder überhaupt keine statistischen Methoden bei der Ergebnisauswertung eingesetzt oder sie wurden zumindest nicht beschrieben.

Bei der Beurteilung der Qualität der Ergebnisdarstellung wurde auch die *kritische Auseinandersetzung* mit den Ergebnissen der eigenen Studie gezählt. In diesem Zusammenhang wurde überprüft, ob die Autoren das eigene methodische Vorgehen reflektiert haben und ob sie sich mit den ethischen Grundlagen ihrer Arbeit auseinandergesetzt haben. Auch soll dargestellt werden, in welchem Umfang die Autoren die Ergebnisse ihrer Studien vor dem Hintergrund der Ergebnisse aus anderen Studien kritisch untersucht und verglichen haben. Dabei ist wichtig, dass nicht allein die Erwähnung anderer Arbeiten als kritische Auseinandersetzung gewertet wurde, sondern dass aus dem Text hervorgehen muss, dass der Autor die eigene Studie tatsächlich mit denen anderer verglichen und Unterschiede zwischen den Studien oder Probleme beleuchtet hat. In dem folgenden Diagramm sind die Ergebnisse dieser Untersuchung zusammengefasst:

⁸⁰ Vgl. Kapitel 2.



Aus der zusammenfassenden Darstellung geht hervor, dass eine allgemeine, kritische Auseinandersetzung der Studienergebnisse in 47 (30 %) von 158 Arbeiten stattgefunden hat.

Die ethischen Grundlagen der Studie wurden in nur 13 Studien reflektiert, was einem Anteil von ca. 8 % entspricht. Zu methodischen Problemen⁸¹ hingegen wurde in 96 (61 %) Studien Stellung bezogen. Es lässt sich aus dieser Darstellung heraus vermuten, dass es einem Großteil der Autoren wichtiger war, die Ergebnisse der Studie darzustellen, als diese kritisch mit den Ergebnissen aus anderen Arbeiten in Bezug zu setzen. Auch wenn nicht unbedingt davon ausgegangen werden kann, dass in den 111 Studien, in denen keine kritische Auseinandersetzung beschrieben wurde, auch tatsächlich keine erfolgt ist, muss man dennoch auf diese Tatsache hinweisen, da sie für eine unkritische Haltung einiger Autoren spricht. Die angewandte methodische Umsetzung der Studien wird dagegen von einem Großteil der Autoren diskutiert. Bei genauer Betrachtung wird deutlich, dass die Diskussion der methodischen Grundlage stark variiert. So wird von vielen Autoren angesprochen, dass sie sich um die Homogenität des Patientenkollektivs mühten. Insbesondere in den Studien, in denen ein alternierendes Verfahren angewendet wurde, wurde auf diese Tatsache besonderen Wert gelegt. Ein anderer von den Autoren ebenfalls häufig erwähnter Aspekt ist die Größe des Patientenkollektivs. Dieser kann in Abhängigkeit vom Kontext sowohl einem methodischen oder auch einem statistischen Problemverständnis zugeordnet werden.

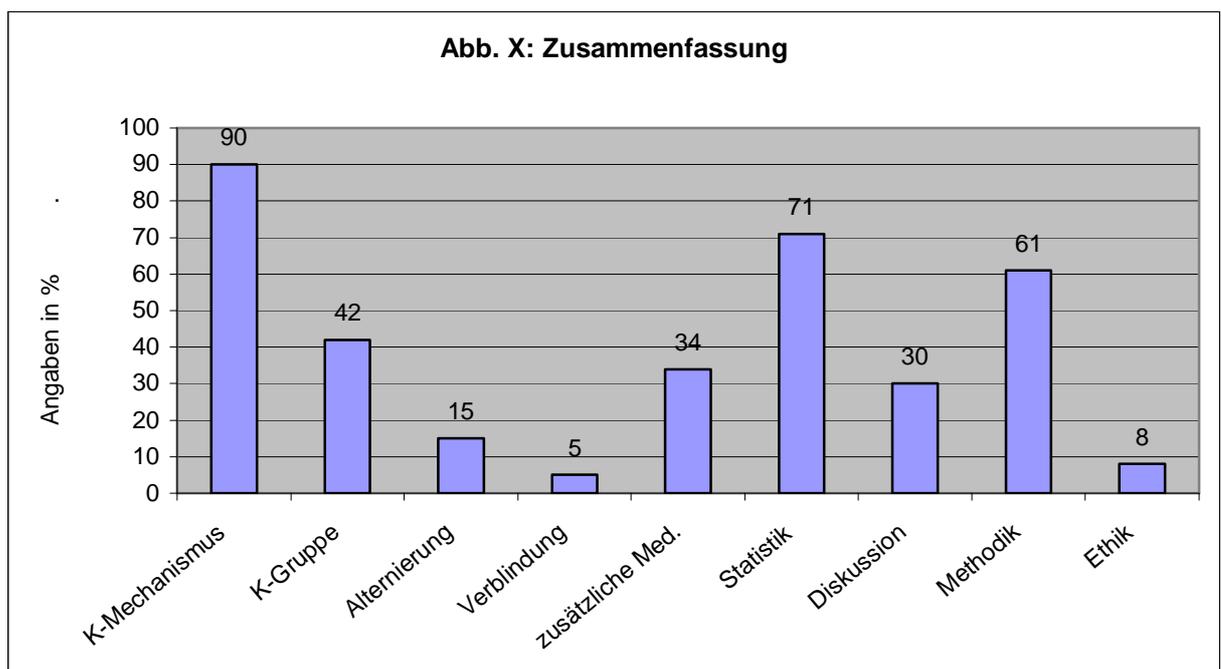
⁸¹Anmerkung: Darunter sind zum Beispiel Stellungnahmen zur mangelnden statistischen Aussagefähigkeit der Ergebnisse zu verstehen.

Ein weiterer Punkt, der insbesondere vor dem Hintergrund der zeitgenössischen politischen Situation brisant ist, ist die Frage, ob von den Autoren *ethische Aspekte* ihrer Arbeit diskutiert wurden. Es finden sich nur wenige Studien, in denen auf die medizinisch-ethische Situation eingegangen wird. Diese Tatsache als Hinweis zu werten, dass die Medizin „unethisch“ gewesen sei, scheint nicht haltbar zu sein. Denn auch in Studien, die den aktuellen Standards klinischer Forschung genügen, wird selten auf ethische Fragestellungen Bezug genommen.

Am Ende dieses Kapitels bietet es sich an, die Ergebnisse zur methodischen Untersuchung der Arbeiten zusammenzufassen, um einen besseren Überblick zu gewähren. Die Übersicht orientiert sich an folgenden Punkten, die auch von Stoll in ihrer Arbeit verwendet wurden:

1. Beschreibung eines Kontrollmechanismus
2. Vorhandensein einer Kontrollgruppe
3. Erwähnung eines verblindeten Versuchsaufbaus
4. Beschreibung eines alternierenden Verfahrens bei Gruppenbildung
5. Erwähnung von Störgrößen
6. Anwendung statistischer Auswertungsverfahren
7. Kritische Diskussion der Ergebnisse
8. Berücksichtigung methodischer Aspekte
9. Bezug auf ethische Fragestellungen

Beschreibt man das vorliegende Material auf der Grundlage dieser Punkte, ergibt sich folgendes Diagramm:



Dem Diagramm ist zu entnehmen, dass ein Kontrollmechanismus in 90 % der untersuchten Veröffentlichungen zu finden war. Kontrollgruppen wurden in 42 % gebildet. Ein alternierendes Verfahren, was Stoll in ihrer Dissertation als „Quasi-Randomisierung“ bezeichnet, konnte in 15 % aller Fälle beziehungsweise in 34 % der mehrarmigen Studien festgestellt werden. Die Einsicht, dass Studienergebnisse überhaupt eine Kontrolle zur Beurteilung brauchen, scheint durchaus weit verbreitet zu sein. Auch die Art des alternierenden Verfahrens zur Gruppenbildung stellt keine Rarität dar, sondern darf als gebräuchlich verstanden werden. Das Bewusstsein für die Erfassung von Störgrößen mittels Beschreibung zusätzlicher Medikamente scheint aus dieser Darstellung nicht so weit verbreitet zu sein. Eine statistische Auswertung der gewonnenen Ergebnisse wird in unterschiedlicher Qualität von 71 % der Autoren beschrieben. Auch das Verständnis für methodische Probleme findet sich in 60 % der evaluierten Arbeiten wieder. Etwas unverständlich erscheint es daher, dass im Kontrast dazu in nur 30 % der Studien eine kritische Auseinandersetzung in Form von Vergleichen mit anderen Arbeiten erfolgt ist. Auf ethische Fragestellungen wird in 8 % der Veröffentlichungen eingegangen. Bei der Beobachtung der Ergebnisse lässt sich der Eindruck gewinnen, dass die klinische Forschung in Deutschland zwei Gesichter zeigte. Auf der einen Seite lässt sich aus der Darstellung verfolgen, dass für die Zeit moderne methodische Herangehensweisen nicht nur von einigen wenigen, sondern wie im Beispiel der alternierenden Gruppenbildung von einer beachtlichen Zahl durchgeführt wurde. Dieser Eindruck bestätigt sich, wenn man die Ergebnisse dieser Arbeit mit denen von Stoll in Beziehung setzt, worauf später eingegangen wird. Andererseits muss man aber auch feststellen, dass ein Bewusstsein für zum Beispiel Störfaktoren und der kritische Umgang mit den Studienergebnissen bei vielen Autoren nicht zu erkennen ist.

3.3. Zeitliche Entwicklung

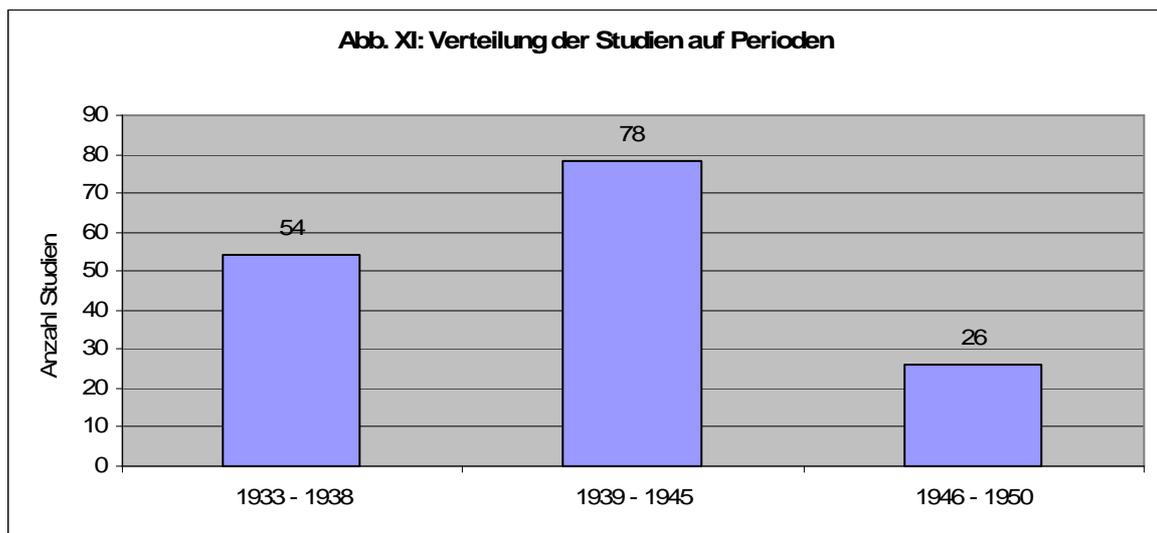
In diesem Abschnitt wird untersucht, ob sich Veränderungen hinsichtlich der Studiengestaltung im zeitlichen Verlauf zeigen. Dazu wurden, um eine übersichtliche Darstellung zu ermöglichen, drei Perioden des untersuchten Gesamtzeitraums gebildet. Die Bildung der Perioden orientiert sich an politischen Ereignissen. Die erste Periode wird definiert als die *Vorkriegsperiode* (1933 – 1938), die zweite als die *Kriegsperiode* (1939 – 1945) und die dritte als die *Nachkriegsperiode* (1946 – 1950).

Zunächst wird untersucht, wie sich die Anzahl der Studien und die einzelnen *Studientypen* – Outcomesstudie, Fallserie, Querschnittsstudie und Experiment - auf die Perioden verteilen. Anschließend werden die wesentlichen Merkmale der *Studiendurchführung* und ihre Umsetzung in den Perioden erfasst und abschließend die Merkmale der Ergebnisdarstellung.

Da sich die Methodik dieser Arbeit stark an der von Stoll anlehnt, bietet sich zudem ein Vergleich der Zeiträume 1918 – 1932 und 1933 – 1950 an.

1. Untersuchung des Zeitraumes zwischen 1933 - 1950

Im ersten Diagramm ist die Verteilung der Studien auf die drei Perioden dargestellt:⁸²



Es zeigt sich, dass in der Vorkriegsperiode etwa 35 %, in der Kriegsperiode etwa 48 % und in der Nachkriegsperiode ca. 17 % der ausgewerteten Studien erschienen sind. Setzt man

⁸² Die Periodenbildung ist auf der Grundlage politischer Ereignisse zustande gekommen, wodurch die Perioden unterschiedlich lange Zeiträume abbilden.

die Studienanzahl in Relation zu der Anzahl der Jahre, dann errechnet sich für die Vorkriegsperiode ein Durchschnittswert von neun Studien pro Jahr, für die Kriegsperiode elf Studien pro Jahr und für die Nachkriegsperiode fünf Studien pro Jahr. Der kleine Anteil der ausgewerteten Studien in der Nachkriegsperiode erklärt sich vor allem als Folge des Nichterscheinens der meisten medizinischen Zeitschriften unmittelbar nach dem Krieg.⁸³ Der Druck der Blätter wurde erst im Lauf der ersten Jahre wieder aufgenommen, so dass man davon ausgehen kann, dass erst in den letzten Jahren der 40er Dekade das Vorkriegsniveau mit einer durchschnittlichen Anzahl von neun Studien jährlich wieder erreicht wurde. Betrachtet man in diesem Zusammenhang die Jahre von 1946 – 1948, so sind in dem Zeitraum von drei Jahren insgesamt nur acht gegenüber 18 Studien in den letzten beiden Jahren 1949 und 50 erschienen.

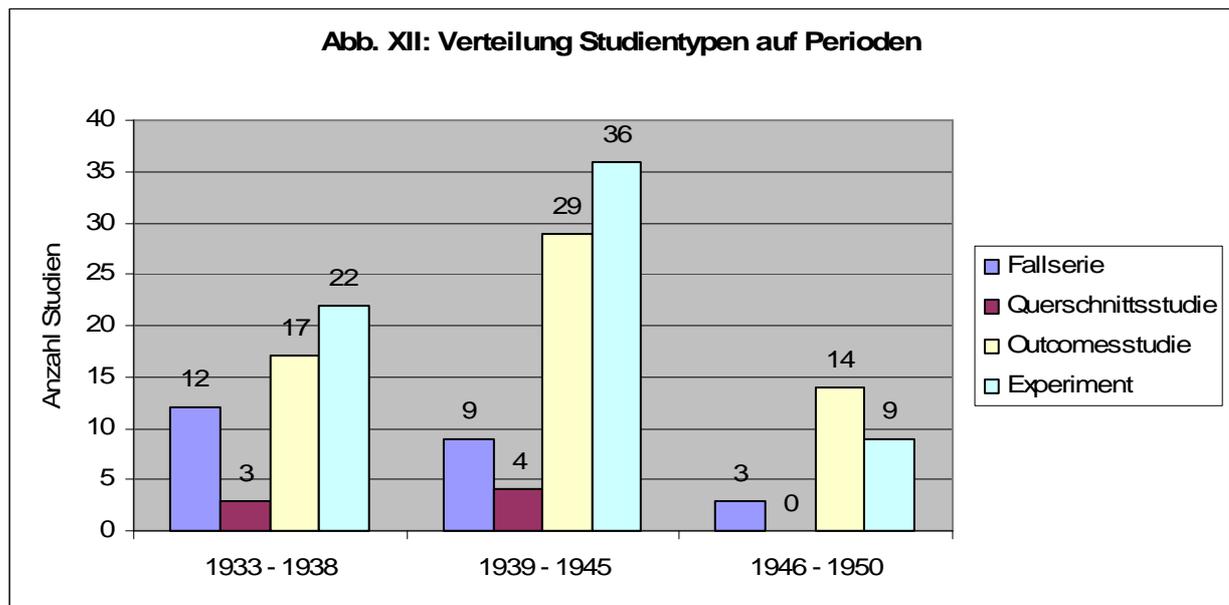
Ein Maß für die **Konzentration** therapeutischer Studien in den medizinischen Fachzeitschriften lässt sich gewinnen, indem man sie in Bezug zur Gesamtzahl aller veröffentlichten größeren Artikel setzt.⁸⁴ Für die Vorkriegsperiode ergibt sich eine durchschnittliche Anzahl von ca. 2000 Publikationen jährlich, für die Kriegsperiode ca. 1000 und für die Nachkriegsperiode etwa 700.⁸⁵ Setzt man die Anzahl der ausgewerteten therapeutischen Studien in Bezug zu der Gesamtzahl, so ergibt sich für die Vorkriegsperiode ein Wert von weniger als **0,5 %** ($9/2000 = 0,0045$), für die Kriegsperiode ein Wert von etwa **1 %** ($11/1000 = 0,011$) und für die Nachkriegszeit ein Wert von ca. **0,7 %** ($5/700 = 0,007$). Fällt beim Vergleich der absoluten Zahlen der Unterschied zwischen der Vor- und der Kriegsperiode (neun Studien in Vorkriegsperiode / elf Studien in Kriegsperiode) gering aus, zeigt sich bei Berücksichtigung aller Veröffentlichungen, dass sich, die Konzentration therapeutischer Studien in diesem Zeitraum annähernd verdoppelt hat. In der Nachkriegsperiode nimmt der Anteil therapeutischer Studien am Gesamtmaterial gegenüber der Kriegsperiode ab, bleibt aber deutlich über dem der Vorkriegsperiode.

Vergleicht man die Verteilung der **Studientypen** auf die Perioden, zeigt sich folgendes Bild:

⁸³ Die Deutsche Medizinische Wochenschrift ist 1945, Die Münchner Medizinische Wochenschrift ist zwischen 1945 und 1950, die klinische Wochenschrift ist zwischen 1945 und 1947 nicht erschienen.

⁸⁴ Anmerkung: Kleine Artikel mit weniger als 50 Zeilen, Zuschriften von Lesern, persönliche Widmungen oder Anmerkungen wurden von der Zählung ausgeschlossen.

⁸⁵ Diese Zahlen stellen eine Schätzung dar, die sich auf eine genaue Durchsicht der einzelnen Zeitschriftenbände gründet.

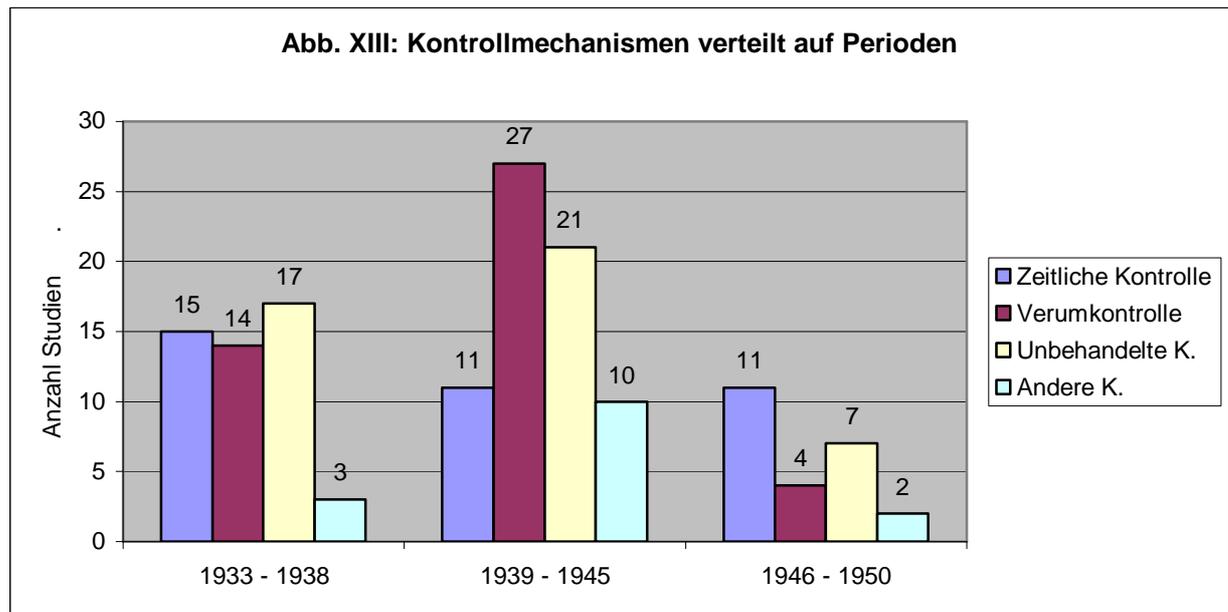


Der Anteil experimenteller Studien in der Vorkriegsperiode liegt in relativen Zahlen ausgedrückt bei ungefähr 41 %, der von Outcomesstudien bei etwa 31 %, der Anteil der Fallserien bei ca. 22 % und der von Querschnittstudien bei knapp sechs %. Zum Vergleich zeigt sich in der Kriegsperiode folgende Verteilung: 44 % experimentelle Studien, 39 % Outcomesstudien, 12 % Fallserien und 5 % Querschnittstudien. In der Nachkriegsperiode ließen sich knapp 54 % der Studien als Outcomesstudie identifizieren, während experimentelle Studien mit ca. 35 % vertreten sind. Der Anteil der Fallserien bleibt gegenüber der Kriegsperiode mit einem Anteil von 11 % konstant. Unter den 26 in der Nachkriegsperiode ausgewerteten Studien fand sich keine Studie mit einem Design einer Querschnittsstudie. Es fällt vor allem auf, dass zwischen der Vor- und der Kriegsperiode insbesondere der Anteil der Fallserien deutlich um ca. 10 % abnimmt, während zwischen der Kriegs- und der Nachkriegsperiode der Anteil experimenteller Studien stark zugunsten des Anteils der Outcomesstudien schrumpft. Kann der Anteil der Querschnittstudien zwischen der Vor- und der Kriegsperiode noch als konstant betrachtet werden, verschwinden diese in der Nachkriegsperiode völlig.

Die Frage nach der **Studiendurchführung** wurde an verschiedenen Studienmerkmalen untersucht.

Zunächst wurden die angewandten **Kontrollmechanismen** untersucht. Hierbei wurden die Kontrollmechanismen anders als im vorherigen Teil zu vier Gruppen zusammengefasst. In der ersten Gruppe wurden die Studien mit einer zeitlichen Kontrolle erfasst, in der zweiten die Arbeiten, in denen eine Verumkontrolle durchgeführt wurde, und in der dritten

diejenigen Veröffentlichungen mit einer unbehandelten Kontrolle. Unter der Kategorie zeitliche Kontrolle wurden die Veröffentlichungen zusammengefasst, in denen zum Beispiel ein historischer Vergleich oder auch Vor- und Nachbeobachtungskontrollen durchgeführt worden sind. Die Studien, in denen ein anderer Kontrollmechanismus eingesetzt wurde, wurden in der vierten Gruppe zusammengefasst. Die 16 Studien, in denen kein Kontrollmechanismus erwähnt wurde, werden in dem folgenden Diagramm nicht aufgeführt:



In der Vorkriegsperiode werden die Mechanismen der zeitlichen, der Verum- und der unbehandelten Kontrolle in einem vergleichbar großen Maßstab eingesetzt. Im Vergleich zur Kriegsperiode fällt vor allem der sprunghafte Anstieg der Verumkontrolle auf. Ihr Anteil beträgt in der Vorkriegsperiode noch 26 % und wächst in der Kriegsperiode auf 35 % an. Der Anteil der zeitlichen Kontrolle sinkt anteilmäßig von 28 % auf 15 %. In der Nachkriegsperiode steigt der Anteil der Studien mit einer zeitlichen Kontrolle auf über 40 % an, während vor allem der Anteil der Studien mit Verumkontrolle um 15 % schrumpft. Relativ konstant verhält sich der Anteil von Studien mit unbehandelten Kontrollen. In der Vorkriegsperiode finden sich fünf Studien (9 %) ohne Kontrollmechanismus, in der Kriegsperiode neun (12 %) und in der Nachkriegsperiode zwei (8 %).

Bei der Frage nach der Studiendurchführung wurde weiter untersucht, ob in den Perioden eine *Vorausplanung* der Studien fassbar ist.⁸⁶

⁸⁶ Anmerkung: Bei dem Merkmal „Vorausplanung“ wurde untersucht, inwieweit zum Beispiel Rekrutierungskriterien existierten.

In der Vorkriegsperiode wurden in 35 (65 %) Studien keine Angaben zur Rekrutierung von Patienten gemacht, und nur in 19 Studien (35 %) wurden solche Kriterien beschrieben. In der Kriegsperiode zeigt sich ein anderes Bild. In diesem Zeitraum wurde die Anwendung von Rekrutierungskriterien in 37 von 75 (49 %) Publikationen erwähnt. In der Nachkriegsperiode beträgt die Anzahl der Studien mit definierten Kriterien 12/26, das entspricht einem Anteil von etwa 46 % entspricht.

Für den Bereich der Studiendurchführung wurde weiter untersucht, ob auf eine **gleichwertige Behandlung** der Patienten geachtet wurde und ob sich zwischen den Perioden Unterschiede zeigen. In der Vorkriegsperiode geht aus dem größten Teil der Studien - etwa 56 % - nicht sicher hervor, ob eine sonst gleichwertige Behandlung der Patienten durchgeführt wurde. Ab 1939 wird in der überwiegenden Mehrheit der Arbeiten - 57 % - eine Gleichbehandlung der Patienten beschrieben. In der Nachkriegsperiode wird auf dieses Kriterium in 65 % der Veröffentlichungen hingewiesen. Eine Ungleichbehandlung der Patienten wird in einem über die Perioden hinweg relativ konstanten Verhältnis selten beschrieben.

Ein weiteres die Studienqualität kennzeichnendes Merkmal stellt die Erfassung von **Störgrößen** dar. Hierzu wurde überprüft, ob Nebenwirkungen, die, wie im zweiten Kapitel erläutert, einen störenden Einfluss auf die Durchführung der Studie haben und daher als Störgröße aufgefasst werden können, ob zusätzliche Medikamentengaben beschrieben wurden und ob sich Unterschiede zwischen den Perioden erkennen lassen. In der Vorkriegszeit wurden in 30 (56 %) von 54 Studien Nebenwirkungen beziehungsweise das Nichtauftreten von denselben beschrieben. In der Kriegsperiode vergrößern sich Anzahl und Anteil der Studien, in denen auf die Wahrnehmung von Nebenwirkungen geachtet wurde, auf 50 von 75 Studien also auf etwa 67 %. In der Nachkriegszeit ändert sich dieses Bild und es wird nur noch in elf (42 %) von 26 Veröffentlichungen auf die Erfassung von Nebenwirkungen Wert gelegt.

In den meisten Veröffentlichungen der Vorkriegsperiode lässt sich nicht verfolgen, ob zu dem in der Studie zu prüfenden Mittel andere Medikamente additiv eingesetzt wurden. So finden sich in 38 (70 %) Arbeiten diesbezüglich keine Angaben, während in zwölf (22 %) eine zusätzliche Medikamentengabe beschrieben und nur in vier (7 %) dieselbe verneint wird. In der Zeit zwischen 1939 und 1945 wird in anteilmäßig mehr Publikationen (37 %) die zusätzliche Gabe von Präparaten erwähnt, während der Anteil der Arbeiten, in denen sich keine Angaben darüber befinden, nur noch 56 % beträgt. In ebenfalls etwa 7 % der Fälle wird eine weitere Medikamentengabe ausgeschlossen. In der Nachkriegsperiode

wächst der Anteil der Studien, in denen zusätzlich verabreichte Präparate erwähnt wurden, noch mal auf 42 %, der Teil ohne Angaben diesbezüglich beträgt 54 % und der, in denen eine zusätzliche Medikamentengabe ausgeschlossen wurde, liegt bei etwa 4 %.

Die Frage, wie in den Perioden die Studien *ausgewertet* wurden, wird anhand von zwei Merkmalkomplexen untersucht. Zum einen wird geprüft werden, wie die Ergebnisse *analysiert* wurden und zum anderen wird betrachtet werden, ob sich die Forscher mit *ethischen* oder *methodischen* Problemen der Forschung auseinandergesetzt haben.

Betrachtet man zunächst den Umgang mit *Subgruppen* in den einzelnen Perioden, so zeigen sich zwischen den ersten beiden Perioden keine bemerkenswerten Veränderungen. So wurde zwischen 1933 und 1938 eine Subgruppenanalyse in 28 von 54 Studien durchgeführt. Das entspricht einem Anteil von etwa 52 %. In der Kriegsperiode ändert sich bei einem Anteil von ca. 53 % daran nur wenig. In den Nachkriegsjahren wird dann in 20 von 26 Studien eine Subgruppenanalyse bei der Auswertung eingesetzt, was einem Anteil von ca. 77 % entspricht.

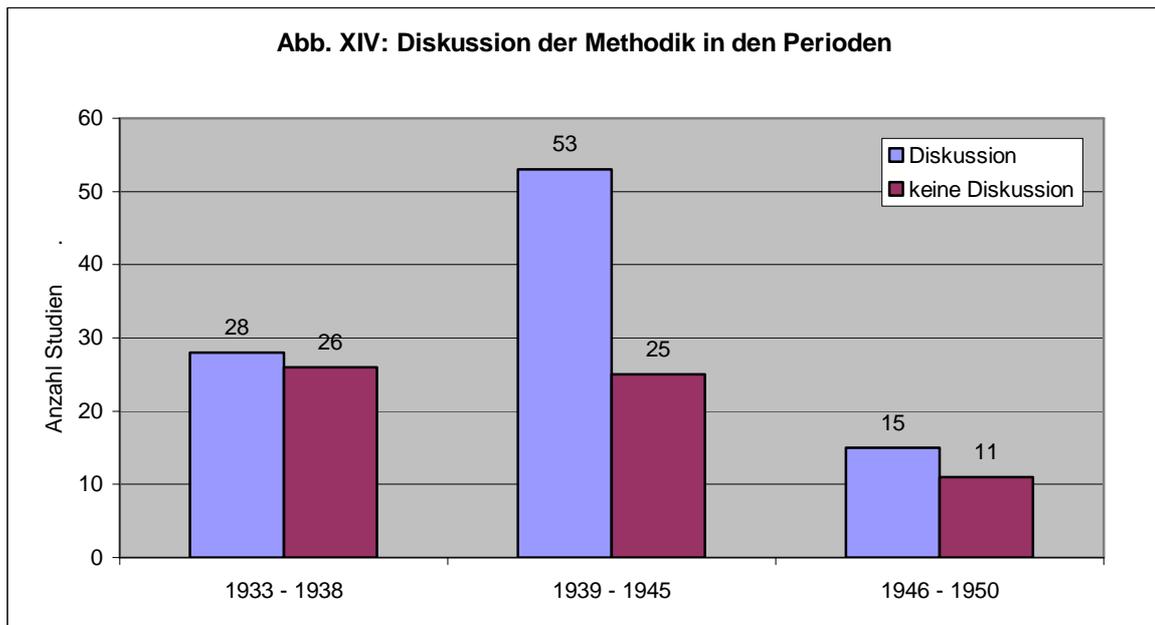
Auch wurde untersucht, ob sich hinsichtlich der *statistischen Auswertung* Veränderungen zwischen den Perioden erkennen lassen. In der Vorkriegsperiode wurden in 34 (62 %) Studien einfache statistische, in drei (6 %) komplexe und in 17 (31 %) Veröffentlichungen keine statistischen Analysen durchgeführt. Zum Vergleich fanden sich in der Kriegsperiode 41 (55 %) Arbeiten mit einer einfachen, zwölf (16 %) mit einer komplexen und 22 (29 %) mit keiner statistischen Auswertung. In der Zeit zwischen 1946 und 1950 wurden in 16 (62 %) Studien einfache, in vier (15 %) komplexe und in sechs (23 %) keine statistischen Methoden angewandt. Der Anteil der Studien mit einer komplexen statistischen Auswertung nahm insbesondere zwischen den ersten beiden Perioden zu und bleibt in der Nachkriegszeit auf dem Niveau der Kriegsperiode.

Eine *kritische Auseinandersetzung* mit der eigenen Arbeit bezieht sich vor allem darauf, ob überhaupt eine zusammenfassende Darstellung der Ergebnisse gegeben wurde und ob die Forscher ihre Resultate mit denen aus anderen Studien verglichen haben.⁸⁷ In der Vorkriegszeit kann für 13 von 54 Veröffentlichungen (24 %) ein *kritischer Umgang* mit den eigenen Ergebnissen festgestellt werden. In dem Zeitraum zwischen 1939 und 1945 wird von 28 Autoren (37 %) eine kritische Beurteilung durchgeführt. In der Nachkriegszeit fällt der Anteil wieder auf das Vorkriegsniveau zurück und so kann nur noch für sechs von

⁸⁷ Anmerkung: Das Merkmal „kritische Auseinandersetzung“ ließ sich valide an der Überprüfung messen, ob die Forschungsergebnisse zusammenfassend dargestellt wurden und ob die eigenen Ergebnisse im Zusammenhang mit dem damaligen Forschungsstand diskutiert wurden.

26 Studien (23 %) eine kritische Auseinandersetzung mit den Ergebnissen nachgewiesen werden.

Die Überprüfung, in welchem Maß sich die Forscher mit *methodischen Gesichtspunkten* der Forschung in den verschiedenen Perioden auseinandergesetzt haben, bietet die Möglichkeit, Veränderungen in dem Bewusstsein für diesbezügliche Fragen zu erkennen.⁸⁸ Das folgende Diagramm zeigt die Verteilung der Studien auf die Perioden, in denen auf methodische Aspekte eingegangen wurde:



In der Vorkriegsperiode kann die Wahrnehmung und Diskussion von methodischen Fragestellungen in etwas mehr als der Hälfte der Arbeiten festgestellt werden. In der Zeit zwischen 1939 und 1945 äußern sich genau 2/3 der Autoren zu methodischen Problemen der Forschung. In der Nachkriegszeit äußern sich absolut und relativ betrachtet wieder deutlich weniger Autoren zu diesen Aspekten, auch wenn der Anteil knapp über dem der Vorkriegsperiode liegt.

Die *ethischen Rahmenbedingungen* der Versuche wurden insgesamt selten in den Studien diskutiert.⁸⁹ In der Vorkriegsperiode wurde in 2 (4 %) von 54 Veröffentlichungen auf ethische Fragestellungen Bezug genommen. In der Kriegsperiode gingen Autoren in neun (12 %) von 75 und in der Nachkriegsperiode in zwei (8 %) von 26 Studien auf ethische Probleme ein.

⁸⁸ Anmerkung: Unter dem Merkmal „methodische Diskussion“ ist zu verstehen, ob der Forscher die Versuchsdurchführung und Auswertung kritisch hinterfragt hat, indem er zum Beispiel die Gruppenbildung oder aber die statistische Aussagefähigkeit seiner Ergebnisse erörtert hat.

⁸⁹ Anmerkung: Studien, in denen zum Beispiel auf eine alternierende Behandlung aufgrund der Überlegenheit eines Mittels verzichtet wurde, wurden in dieser Kategorie zusammengefasst.

Insgesamt ist es schwierig, eine eindeutige Entwicklung oder Tendenz zwischen den Perioden ablesen zu wollen.

Die ungleichmäßige Verteilung der Anzahl der Studien zugunsten vor allem der Kriegsperiode weist auf ein zunehmendes Interesse an klinischer Forschung in dieser Zeit hin. Dies wird vor allem deutlich, wenn man berücksichtigt, dass spätestens ab 1940 die Auswirkungen des Krieges dazu führen, dass insgesamt deutlich weniger Veröffentlichungen in den Zeitschriften abgedruckt werden und dennoch die meisten ausgewerteten Publikationen in diesem Zeitraum zu finden sind. Erst ab 1944 führt die deutliche Reduktion der Gesamtveröffentlichungen zu einer gleichgerichteten starken Abnahme der Anzahl klinischer Studien. Die geringe Anzahl klinischer Studien bis etwa 1948 deutet daraufhin, dass der Krieg nachhaltige Auswirkungen auf die Situation der klinischen Forschung in Deutschland gehabt hat.

Hinsichtlich der angewandten **Kontrollmechanismen** ist bemerkenswert, dass sich zwischen den ersten beiden Perioden eine Entwicklung zeigt, die die Verwendung von Verum- und unbehandelten Kontrollen begünstigt. Doch im Vergleich zur Nachkriegszeit scheint sich dieser Trend umzukehren und es wird wieder verstärkt vor allem auf zeitliche Kontrollmechanismen zurückgegriffen. Für die Studienplanung lässt sich feststellen, dass im Verhältnis die Anwendung von Rekrutierungskriterien in den letzten beiden Perioden eine stärkere Beachtung erfährt als in der Vorkriegsperiode. Auch wird im Vergleich zur Vorkriegsperiode in der Kriegs- und Nachkriegszeit vermehrt auf die Gleichbehandlung von Patienten hingewiesen.

Bei der Erfassung von **Störgrößen** zeigt sich ein uneinheitliches Bild. Wird die Beschreibung von Nebenwirkungen zu untersuchender Präparate in der Kriegszeit relativ betrachtet häufiger durchgeführt als in der Vorkriegszeit, setzt sich diese Entwicklung in der Nachkriegsperiode nicht fort, und es wird sogar anteilmäßig in geringerem Maß als in der Vorkriegsperiode die Erfassung von Nebenwirkungen verfolgt. Bei der Erfassung der Störgröße „*zusätzliches Medikament*“ kann festgestellt werden, dass über die Perioden hinweg der Anteil der Studien, in denen die Gabe von zusätzlichen Medikamenten bewusst genannt wird, zunimmt.

Für die Merkmale der **Studienauswertung** lässt sich ebenfalls keine einheitliche Entwicklung aufzeigen. Die über die Perioden hinweg zunehmende Bildung von Subgruppen zur Auswertung kann aus heutiger Sicht als ein gewisser Fortschritt betrachtet werden. Bei der Anwendung statistischer Merkmale lässt sich ein solcher Trend nicht nachvollziehen. Im Vergleich zwischen den ersten beiden Perioden wird eine komplexere

statistische Auswertung durchaus deutlich häufiger verwendet, aber in der Nachkriegszeit stagniert die Anwendung komplexer statistischer Verfahren auf dem Niveau der Kriegsperiode. Ähnlich verhält es sich mit den Merkmalen der „allgemeinen kritischen Auseinandersetzung“ und der „methodischen Auseinandersetzung“ in den Publikationen. Zusammenfassend kann man feststellen, dass sich in den beobachteten Perioden einige aus heutiger Sicht fortschrittliche Veränderungen hinsichtlich methodischer Aspekte in den Publikationen angedeutet haben. Insbesondere lässt sich das für die Anwendung von Rekrutierungskriterien und für die Beachtung der Gleichbehandlung der Patienten beobachten. Bei einigen anderen Studienmerkmalen lassen sich einige Hinweise erkennen, die insbesondere zwischen der Vor- und der Kriegsperiode eine aus heutiger Sicht positive Entwicklung vermuten lassen. Dass sich dann aber im Vergleich zwischen Kriegs- und Nachkriegsperiode diese Weiterentwicklung häufig nicht verfolgen lässt, zeigt, dass sich der Gebrauch bestimmter methodischer Merkmale scheinbar nicht hat durchsetzen können. Das lässt sich besonders an den Beispielen der Anwendung statistischer Verfahren oder auch für den Bereich von Störgrößen verfolgen.

2. Vergleich der Zeiträume von 1918 - 1932 und 1933 – 1950

Wie eingangs bereits erwähnt, bietet sich ein Vergleich zu der Arbeit von Stoll an, die mit einer vergleichbaren Methode den Zeitraum zwischen 1918 - nach Ende des ersten Weltkrieges - und 1932 – das Erscheinungsjahr der ersten Auflage der Methodenlehre von Martini – untersucht hat.⁹⁰

In ihrer Arbeit hat Stoll 268 klinisch-therapeutische Studien in dem genannten Zeitraum mittels der Checkliste evaluiert. Das ergibt einen Durchschnittswert von etwa 19 Studien pro Jahr. Bei der Typisierung der Studien fand Stoll, dass 49 % der Studien in Form eines Experimentes, 25 % in Form einer Beobachtungsstudie und 18 % in Form einer Fallserie gestaltet wurden. Nur 0,5 % der Arbeiten wurden als Querschnittsstudie typisiert und 3,5 % ließen sich keiner der Kategorien zuordnen.⁹¹

Zum Vergleich wurden in dieser Arbeit bei der gleichen Anzahl von Zeitschriften 158 Studien auf 18 Jahre verteilt ausgewertet, was einem Durchschnittswert von etwa neun Studien pro Jahr entspricht. 42 % der Studien wurden als experimentelle Studien, 39 % als

⁹⁰ Stoll 2005, S. 27 f.

⁹¹ Stoll 2005, S. 99.

Beobachtungsstudien, etwa 15 % als Fallserien und 5 % als Querschnittstudien charakterisiert.

Kontrollmechanismen hat Stoll in 68 % der 268 Studien beobachtet. Am häufigsten wurde von ihr mit 51 % ein Verumvergleich als Kontrollmechanismus beschrieben. Eine zeitliche Kontrolle wurde in insgesamt 34 % und eine unbehandelte Kontrolle in 21 % der Studien angewandt. Placebokontrollierte Versuche wurden in 5 % der Studien beobachtet.⁹²

Für den Zeitraum von 1933 – 1950 war in 90 % der Studien ein Kontrollmechanismus nachweisbar. Der Verumvergleich als Kontrolle ist in knapp 30 %, der Vergleich mit einer unbehandelten Gruppe in 31 % der Publikationen zu erkennen. Eine zeitliche Kontrolle lässt sich in ca. 27 % der Studien beobachten. Placebokontrollierte Versuche wurden, ähnlich wie bei Stoll, in geringem Maß durchgeführt und finden sich in etwa 3 % der Studien.

Bei der Patientenrekrutierung findet Stoll in 47 % der Studien Einschlusskriterien erwähnt. Ausschlusskriterien wurden mit 20 % deutlich geringer angegeben.⁹³ Für die Zeit nach 1932 zeigt sich, dass in 43 % der Arbeiten über den Gebrauch von Einschlusskriterien und in 37 % von Ausschlusskriterien berichtet wurde.

Für experimentelle Studien gibt Stoll an, dass bei 2 % ein alternierendes Verfahren für die Gruppenbildung angewandt wurde.⁹⁴ Eine Verblindung der Studienteilnehmer wird in 5 % der Studien beobachtet.⁹⁵ In dem Zeitraum zwischen 1933 und 1950 finden sich 23 Studien mit einer alternierenden Zuweisung zu den Gruppen, was einem Anteil von 15 % entspricht. Eine Verblindung der Teilnehmer konnte in diesem Zeitraum ebenfalls in 5 % aller Fälle festgestellt werden.

Was den Bereich der Studienauswertung betrifft, stellt Stoll fest, dass in 84 % der Studien eine Operationalisierung der Outcomes, in 67 % eine einfache und in 2 % eine komplexe statistische Auswertung durchgeführt wurde. Eine Subgruppenanalyse erfolgte in 47 % der Studien und in 54 % wurden Dropouts erwähnt.⁹⁶ Demgegenüber wurde in dem Material dieser Arbeit in 96 % der Studien eine Operationalisierung des Outcomes festgelegt, in 58 % wurde eine einfache und in etwa 12 % eine komplexe statistische Auswertung angewandt. Die Untersuchung von Subgruppen fand in ca. 57 % der Arbeiten statt und Dropouts wurden von 27 % der Autoren beschrieben.

⁹² Stoll 2005, S. 100.

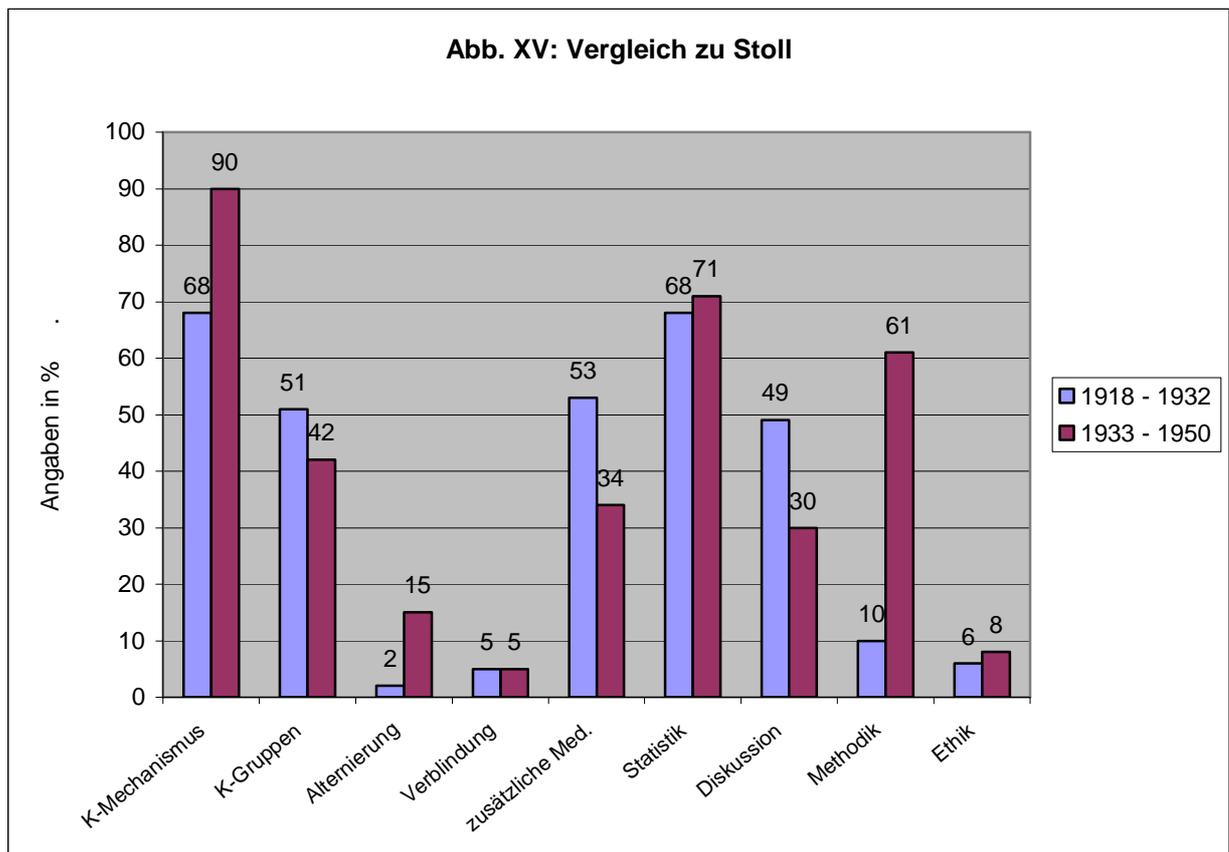
⁹³ Stoll 2005, S. 102.

⁹⁴ Anmerkung: Stoll bezeichnet das alternierende Verfahren als „Quasirandomisierung“; Stoll 2005, S. 106.

⁹⁵ Stoll 2005, S. 106.

⁹⁶ Stoll 2005, S. 111.

In 49 % der Studien erkennt Stoll eine kritische Diskussion der Ergebnisse. Bei 10 % der Arbeiten beobachtet sie eine methodische und bei 6 % eine ethische Auseinandersetzung.⁹⁷ Eine kritische Beurteilung lässt sich in meinem Material in etwa 30 % erkennen. In 60 % der Arbeiten kann man eine methodische und in 8 % eine ethische Auseinandersetzung verfolgen. In folgendem Diagramm werden die wichtigsten Merkmale im Vergleich zu Stoll zusammengefasst:⁹⁸



An der Übersicht lässt sich erkennen, dass die Ergebnisse in einigen Merkmalen weitgehend übereinstimmen, sich bei anderen allerdings zum Teil erhebliche Unterschiede zeigen. Bezüglich der Merkmale „*Verblindung*“ und „*ethischer Diskussion*“ deuten sich keine oder nur geringfügige Unterschiede. Auf den ersten Blick scheint auch die Anwendung statistischer Methoden auf keine deutlichen Veränderungen hinzuweisen. Wenn man aber den Gebrauch komplexer statistischer Mittel miteinander vergleicht, dann ergibt sich ein anderes Bild. Deutliche Unterschiede lassen sich auch in der Beschreibung von Kontrollmechanismen, von Kontrollgruppen und in der Erfassung von Störgrößen in Form von Medikamenten beobachten. Während Kontrollmechanismen scheinbar zunehmend als Grundlage für die Beurteilung eingesetzt werden, lässt sich das für den

⁹⁷ Stoll 2005, S. 112.

⁹⁸ Vgl. dazu Stoll 2005, S. 114.

Einsatz von externen Kontrollgruppen oder die Erfassung von Störgrößen nicht behaupten, und im Gegenteil werden diese methodischen Merkmale seltener als in der Zeit vor 1933 eingesetzt. Besonders auffällig sind die Unterschiede für die Merkmale der methodischen Diskussion und für den Gebrauch des alternierenden Verfahrens. Hier zeigt sich, dass nach 1932 alternierende Verfahren immer häufiger eingesetzt werden. Dass häufiger methodische Fragen erörtert werden, steht scheinbar im Widerspruch zu einigen vorher gemachten Beobachtungen, die eher eine Art Rückschritt der methodischen Umsetzung dokumentieren. Dieser Widerspruch deckt auch gleichzeitig die Grenzen dieses Vergleichs auf. So lassen sich speziell für die Merkmale, die sich mit der Ergebnisdarstellung der kritischen und methodischen Diskussion befassen, schwerer objektivierbare Kriterien formulieren als für andere Kriterien, wie zum Beispiel für die Frage nach einer alternierenden Versuchsanordnung. Beurteilungsmaßstäbe, welche Bedingungen erfüllt sein müssen, um zum Beispiel dem Merkmal einer kritischen Diskussion zu entsprechen, lassen sich schwer standardisieren. Insofern sind die Ergebnisse in einigen Bereichen einer solchen Untersuchung stärker vom Betrachter abhängig als in anderen. Das erhöht das Risiko für eine objektive Unterscheidung in der Beurteilung. Dennoch halte ich einen Vergleich dieser Untersuchungen für vertretbar, da das Problem unterschiedlicher Bewertungen zwischen zwei Arbeiten nicht für alle Merkmalsbereiche zutrifft.

4. Zur Wahrnehmung von Paul Martini und seiner Methodenlehre

In diesem Teil werden die Ergebnisse aus der Evaluation mit dem verglichen, was Paul Martini bezüglich der Umsetzung klinischer Forschung in seiner Methodenlehre kritisiert und gefordert hat. Dabei werden zunächst die zentralen Punkte dargestellt, die von Martini kritisiert und gefordert werden. Anschließend wird anhand des Materials überprüft, in welchem Umfang Martini wahrgenommen wurde und welche seiner Forderungen berücksichtigt wurden. Zuletzt wird auf eine Diskussion über Martinis Vorschläge hinsichtlich der statistischen Auswertung eingegangen.

1. Zu den Vorschlägen Paul Martinis in der Methodenlehre

Im Vorwort der 1. Auflage des Buches „Methodenlehre der therapeutischen Untersuchung“, welches 1932 veröffentlicht wurde, schreibt Martini:

„Drei Jahre, in denen ich das Glück hatte, die große medizinische Abteilung des St. – Hedwig-Krankenhauses in Berlin zu leiten, waren erfüllt von dem Suchen nach befriedigenden Methoden therapeutischer Untersuchungen. Die vorliegenden Ergebnisse sind noch nichts Vollendetes. Ich veröffentliche sie trotzdem, denn die Änderung des bisherigen Zustandes scheint mir keinen Aufschub zu ertragen.“⁹⁹

13 Jahre später 1945 schreibt Martini in dem Vorwort zur 2. Auflage, die erst 1947 unter dem Titel „Methodenlehre der therapeutisch-klinischen Forschung“ erschienen ist:

„Wer eine besondere Methodenlehre der therapeutisch-klinischen Forschung für nötig erklärt, hält offenbar die einfache ärztliche Erfahrung für unzureichend. Man wendet ein: Gute Ärzte haben in ihrem Leben doch so viel erfahren, haben gründlich beobachtet und haben ihre Meinungen immer wieder kritisch korrigiert! Warum ist dann trotz aller dieser wirklichen oder scheinbaren Garantien die therapeutische Quintessenz der ärztlichen Erfahrung in weiten Bereichen und in vielen Einzelfragen ganz und gar widerspruchsvoll? Für diese Dissonanz gibt es nur die eine Erklärung, dass die gewöhnliche, „naive“ Erfahrung schlechthin und dass auch die bisher geläufigen Methoden der klinischen Forschung den besonderen Ansprüchen therapeutisch-klinischer Probleme nicht genügen.“

⁹⁹ Martini 1932, im Vorwort.

Nach 13 Jahren weiterer Erfahrung hofft das vorliegende Buch diese Lücke vollkommener schließen zu können, als es seinem Vorgänger, der „Methodenlehre der therapeutischen Untersuchung“ 1932 möglich war.“¹⁰⁰

Aus dem Vorwort der Auflage von 1932 geht hervor, dass Martini mit der Veröffentlichung der Methodenlehre die Hoffnung verband, eine Veränderung der Umsetzung klinischer Forschung zu bewirken.

In dem Vorwort zur 2. Auflage kritisiert Martini die „naive Erfahrung“ forschender Kollegen und zeigt sich enttäuscht über den Zustand der therapeutischen Forschung und über die Resonanz der 1. Auflage seines Buchs.

Die folgende Darstellung der Empfehlungen Martinis an die klinische Forschung bezieht sich zunächst auf die 1. Auflage der Methodenlehre von 1932.

Im Kapitel „I. Der Vergleich als Grundlage der therapeutischen Untersuchung“ erläutert Martini die Notwendigkeit, einen Vergleich als Maß zur Erfolgsbeurteilung einer Therapie heranzuziehen. Martini definiert Merkmale, die helfen sollen, eine Vergleichsgrundlage zu bestimmen:

„Die Vergleichsgrundlagen können gewonnen werden aus der Beobachtung des Verlaufs einer Erkrankung, ihrer Dauer und ihres Ausgangs; für alle drei muss bekannt sein ihr Verhalten einerseits ohne spezifische Behandlung, andererseits unter solchen.“¹⁰¹

Die Anwendung der erwähnten Merkmale ist seiner Ansicht nach von der Chronizität der Erkrankung abhängig:

„Wir werden sehen, dass diese Verschiedenheit zwischen kurz dauernden akuten und lang dauernden chronischen Erkrankungen weitgehende Konsequenzen für die Art der vergleichenden Untersuchung mit sich bringt.

Wollen wir eine Vergleichsgrundlage für die Beobachtung des Verlaufs einer Erkrankung gewinnen, so wird die ärztliche Beobachtung des einzelnen Kranken vor dem Einsatz des zu prüfenden Heilverfahrens fast immer unentbehrlich sein. Dies ist, wie oben auseinandergesetzt, die Methode der Wahl bei allen chronischen Erkrankungen; Die Vergleichsgrundlage wird erworben mittels einer Vor- oder einer (seltener) Nachbeobachtung; sie muss in jedem Fall der Beobachtungszeit des zu prüfenden Heilmittels benachbart sein.“¹⁰²

Für akut verlaufende Erkrankungen schlägt Martini ein anderes Vorgehen vor:

¹⁰⁰ Martini 1947, im Vorwort.

¹⁰¹ Martini 1932, S. 1.

¹⁰² Martini 1932, S. 2.

„Ist der Verlauf einer Erkrankung zu kurz, als dass eine vergleichende Beobachtung in genügend langen Perioden zustande käme, oder kommt es, wie bei den akuten Erkrankungen gewöhnlich, überhaupt nicht zu einer Kontinuität des Verlaufs, so entfällt die Möglichkeit einer eigenen Vorbeobachtung, und es muss eine andere Vergleichsgrundlage gesucht werden. Diese Möglichkeit ist gegeben bei einem Krankheitsverlauf, der nicht gerade typisch zu sein braucht, aber immerhin so beschaffen sein muss, dass er erlaubt, aus vielen Fällen der gleichen Erkrankung Durchschnittszahlen zu gewinnen für die Dauer, wenigstens aber für den voraussichtlichen Ausgang eines Falls der gleichen Erkrankung unter bestimmten Bedingungen.“¹⁰³

Für Martini ist es entscheidend, den zeitlichen Verlauf und den Ausgang als Vergleichsgrundlage zu berücksichtigen. Weiter beschreibt er, dass die Einteilung in akute oder chronische Erkrankungen maßgeblichen Einfluss auf die Anwendung von Merkmalen als Vergleichsgrundlage hat. Daraus ergibt sich nach Martinis Ansicht, dass bei chronischen Erkrankungen der Vergleich in der Betrachtung des Krankheitsverlaufs beim Individuum erfolgt, während bei akut verlaufenden Erkrankungen der interindividuelle Vergleich die Beurteilungsgrundlage darstellt.

Weiter stellt Martini seine Vorschläge zur methodischen Gestaltung von Untersuchungen an akut verlaufenden Erkrankungen dar:

„Bei starken epidemiologischen Schwankungen wird es überhaupt unmöglich, Krankengruppen aus verschiedenen zeitlichen Perioden miteinander zu vergleichen; die Literatur der Therapie der akuten epidemischen Krankheiten mit ihrer Unzahl von verfrühten Siegesmeldungen ist eine fortlaufende Beweiskette hierfür.

Solchen aus epidemiologischen und anderen Schwankungen entspringenden Schwierigkeiten beugt in souveräner Weise vor die von WAGNER-JAUREGG so benannte ‚*Simultanmethode*‘. Von Patienten gleicher Erkrankung, die unter sonst gleichen Bedingungen stehen, wird abwechselnd einer ohne jedes spezifische Mittel behandelt oder höchstens mit Mitteln, die nach Art und Grad ihrer Wirkung genügend bekannt sind, während dem anderen außerdem das zu prüfende Mittel zugelegt wird; unentbehrliche Mittel (z. B. Analeptica) müssen beiden in gleicher Weise und Menge verabreicht werden.“¹⁰⁴

Zusammengefasst schlägt Martini also vor, dass bei der Erfolgsbeurteilung der *zeitliche Verlauf*, die Betrachtung von *Vor- und Nachperioden* und der *Ausgang* geeignete

¹⁰³ Martini 1932, S. 4.

¹⁰⁴ Martini 1932, S. 4.

Merkmale seien, um den Erfolg einer therapeutischen Maßnahme zu überprüfen. Für chronisch verlaufende Erkrankungen empfiehlt er, Perioden zu bilden, in denen abwechselnd eine rein symptomatische gefolgt von einer therapeutischen Behandlung erfolgt. Auf der Grundlage des Vergleichs dieser Perioden könne dann die Wirksamkeit einer Therapie ermittelt werden. Bei der Beurteilung der Behandlung akuter Erkrankungen empfiehlt Martini ein anderes Vorgehen, welches er die „*Simultanmethode nach Wagner-Jauregg*“ nennt. Dieses Vorgehen ist gleichbedeutend mit einem alternierenden Verfahren.

Im Kapitel „II. Die Mitursachen in der therapeutischen Forschung“ geht Martini zunächst auf die „auszuschaltenden Mitursachen“ ein. Er definiert Mitursachen wie folgt:

„Der eindeutigen Fragestellung nähern wir uns durch die Ausschaltung von *Mitursachen*, die den ‚reinen Fall‘ komplizieren, indem sie entweder das Wesen und den Verlauf der Erkrankung bzw. den Zustand des Kranken in wesentlicher Weise ändern und die Deutung ihrer Symptome unmöglich machen, oder indem sie unser therapeutisches Vorgehen behindern, ihm entgegenarbeiten oder auch es verstärken, jedenfalls indem sie es ändern.“¹⁰⁵

Martini unterscheidet zwischen seelischen und körperlichen Mitursachen:

„Die Mitursachen können seelischer und körperlicher Natur sein. Zu den ersteren gehört die Suggestion, von der ich soeben gesprochen habe; aber auch jede andere bewusste oder unbewusste *psychische Alteration*; jeder Affekt, Freude, Schmerz und Trauer können den Verlauf einer Erkrankung ungünstig oder auch günstig beeinflussen.“¹⁰⁶

Ferner differenziert er noch zwischen vermeidbaren und unvermeidbaren Ursachen. Als unvermeidbare Mitursachen bezeichnet er unter anderem hohes Lebensalter, besondere Körperschwäche, verschiedene komplizierende Begleiterkrankungen und epidemiologische Schwankungen. Unter den vermeidbaren Mitursachen versteht er zum Beispiel zusätzlich verabreichte Medikamente:

„Viel zahlreicher und größtenteils *vermeidbar* sind die Komplikationen (Mitursachen), die *durch Krankenbehandlung und Krankenpflege* willkürlich oder unwillkürlich (versehentlich) in die Versuchsanordnung hineingetragen werden. Der untersuchende Arzt schafft sich selbst Komplikationen, indem er zwei oder gar mehrere gleichgerichtete Heilmittel oder –methoden gleichzeitig anwendet. Eine solche Konkurrenz ist keineswegs

¹⁰⁵ Martini 1932, S. 10.

¹⁰⁶ Martini 1932, S. 10.

immer leicht erkennbar und spielt auch in sonst guten therapeutischen Arbeiten eine verborgene aber umso gefährlichere Rolle.“¹⁰⁷

Er erwähnt drei praktische Empfehlungen zur Ausschaltung von Mitursachen. Um die Wirksamkeit einem Präparat eindeutig zuweisen zu können, schreibt Martini:

„Es ist auch tatsächlich etwas Alltägliches, dass es sich herausstellt, dass die Eindeutigkeit einer Versuchsanordnung ohne Verzicht auf ein ärztlich im gegebenen Fall indiziertes und notwendiges Mittel nicht durchgeführt werden kann; aus solcher Situation gibt es keinen anderen ärztlichen Ausweg, als den Patienten nach ärztlichem Wissen und Gewissen zu behandeln, ihn aber aus der wissenschaftlichen Untersuchung auszuschalten.“¹⁰⁸

Zum einen lässt sich aus dieser Textstelle erkennen, dass Martini empfiehlt, Patienten, die aus ärztlicher Notwendigkeit mit verschiedenen Präparaten behandelt werden müssen, ex post auszuschließen, also so genannte „*Drop-out-Kriterien*“ anzuwenden. Zum anderen ist darin implizit die Forderung enthalten, nach Möglichkeit therapeutische Maßnahmen von anderen isoliert zu untersuchen.

Bereits im ersten Kapitel geht Martini auf die unwissentliche Versuchsanordnung als weitere Maßnahme zur Mitursachenausschaltung ein:

„Das beste Mittel zur Ausschaltung suggestiver oder sonstiger unsachlicher und deshalb unerwünschter Faktoren ist eine *unwissentliche Versuchsanordnung*. Auf die Hauptgruppe unseres therapeutischen Rüstzeugs angewandt heißt das: die *Medikamente* müssen in einer Form oder Umhüllung dem Kranken gegeben werden, dass ihr spezieller Charakter nicht erkannt werden kann, sie müssen *getarnt* werden.(...) Zweck der Tarnung ist in beiden Fällen, den Kranken in Unkenntnis über den Beginn der Verabreichung des zu prüfenden Heilmittels, evtl. auch über die Tatsache der Verabreichung überhaupt zu lassen.“¹⁰⁹

Zusammengefasst umfasst Martinis Maßnahmenpaket zur Ausschaltung von Mitursachen die *unwissentliche Versuchsanordnung*, die kompromisslose Anwendung von „*Drop-out-Kriterien*“¹¹⁰ und die *isolierte Prüfung* jedes einzelnen Medikamentes zur Versuchsanordnung.

Der statistischen Auswertung klinischer Ergebnisse kommt in der Methodenlehre von Martini eine zentrale Rolle zu. Martini leitet das Kapitel „IV. Statistische Methodik“ mit folgenden Worten ein:

¹⁰⁷ Martini 1932, S. 11.

¹⁰⁸ Martini 1932, S. 12.

¹⁰⁹ Martini 1932, S. 9.

¹¹⁰ Anmerkung: Der Begriff „Drop out“ wird von Martini selbst nicht verwendet.

„Wenn es richtig ist, dass der *reine Fall* unerreichbares Ziel der Klinik ist, dann sind die Methoden ihrer Untersuchung notwendigerweise statistische. Das *eindeutig bestimmte Experiment* und die *exakte Induktion* auf der einen Seite, die *statistische Masse* und die *generalisierende Induktion* auf der anderen Seite, das sind die Pole, zwischen denen die therapeutische Forschung sich bewegt. Sie kann die Statistik und deren Gesetze nicht entbehren, weil sie den reinen Fall nie erreichen kann.“¹¹¹

Bei der statistischen Auswertung müssen nach Martini verschiedene Grundsätze berücksichtigt werden. Es sei dem „*Gesetz der großen Zahl*“ nachzukommen, welches mit dem „*Gesetz der Gleichartigkeit*“ konkurriere:

„Offenbar widerstreiten sich diese drei Forderungen: der Grundsatz der Gleichartigkeit zielt auf eine Vielheit der Gruppen und führt so leicht zu Unübersichtlichkeit und Kleinheit der Zahl; umgekehrt gefährdet das Beharren auf großen Zahlen und auf größerer Übersichtlichkeit die Gleichartigkeit. Es ist Sache der Kritik und Erfahrung, dass der Bearbeiter hier den richtigen Mittelweg findet.“¹¹²

Nachdem Martini die Notwendigkeit der statistischen Beurteilung von klinischen Ergebnissen formuliert hat, geht er auf wichtige statistische Größen ein. Dabei nennt er das *arithmetische Mittel*, den *mittleren Wert*, den *häufigsten Wert* und die *statistische Streuung*. Im darauf folgenden Kapitel „V. Die Beurteilung therapeutischer Resultate“ beschreibt er Beispiele für die Auswertung von Versuchsergebnissen. Dabei spielt die *spezifische Wahrscheinlichkeitszahl* eine große Bedeutung in Martinis Ausführungen:

„Übertragen wir die mathematische Übung auf die Bedingungen der therapeutischen Probleme, so ist deren vergleichender Charakter von vorneherein grundlegend zu berücksichtigen. Er hat zur Folge, dass sich das Ziel der mathematischen Behandlung hier als Urteil über die Überlegenheit einer Wahrscheinlichkeit über eine andere herausstellt. Diese Überlegenheit ist nichts anderes als die Differenz der beiden zu vergleichenden Wahrscheinlichkeiten; die Differenz ist die *spezifische Wahrscheinlichkeitszahl* einer Behandlungsmethode. Der einfachste Fall liegt vor bei der *Beurteilung* einer Erkrankung bzw. *eines Heilerfolgs nach dem Ausgang*, ob zur Heilung, ob zum Tode. Die beiden Vergleichspartner, die Gruppen einerseits der unspezifisch und andererseits der spezifisch behandelten Fälle, besitzen ihre gesonderten Wahrscheinlichkeiten, deren Differenz die spezifische Wahrscheinlichkeitszahl ergibt.“¹¹³

¹¹¹ Martini 1932, S. 16 f.

¹¹² Martini 1932, S. 17 f.

¹¹³ Martini 1932, S. 24 f.

Eine genügende statistische Auswertung klinischer Studien umfasst also Martini zufolge eine Reihe von mathematischen Operationen. Zu diesen zählt insbesondere die Berechnung von *arithmetischem Mittel*, *häufigstem Wert* und des *Zentralwerts*. Der *spezifischen Wahrscheinlichkeitszahl* kommt beim Vergleich zweier verschieden behandelter Gruppen eine wichtige Bedeutung zu. Zur Überprüfung der mathematischen Aussagekraft verschiedener statistischer Werte wie zum Beispiel des arithmetischen Mittels empfiehlt Martini die Berechnung der *statistischen Streuung*. Außerdem müsse in der Auswertung die Homogenität des Kollektivs durch Bildung von *Subgruppen* gesichert werden.

In dem Text finden sich verschiedene Hinweise zur Entstehungsgeschichte der Methodenlehre. Dabei wird deutlich, von welchen Ideen und Personen Martini beeinflusst wurde. Einen bedeutenden Einfluss scheint Eugen Bleuler auf Martini gehabt zu haben. Bleuler war Schweizer Psychiater, der 1919 in dem Buch „Das autistisch undisziplinierte Denken in der Medizin und seine Überwindung“ die methodische Durchführung und die Ergebnisauswertung therapeutischer Studien kritisiert. Aus diesem Buch zitiert Martini: „Wenn BLEULER bekennt: ‚Ich war lange in einer Anstalt mit über 700 Patienten, bin über 20 Jahre in einer Klinik mit 400 Betten und einem Wechsel von 200 – 700 Patienten und habe mir recht viel Mühe gegeben, über bestimmt Arzneimittel ein Urteil zu bekommen; abgesehen von ganz wenigen Fällen bin ich aber nicht einmal so weit gekommen, mir eine vorläufige Meinung, geschweige ein abschließendes Urteil zu bilden‘, so sollte dies auch den optimistischsten Therapeuten eine dringende Warnung zur kritischen Selbstbesinnung sein.“¹¹⁴

Es finden sich in dem Text noch weitere Verweise auf Bleuler und es scheint, dass dieser für Martini ein wichtiger Impuls- und Ideengeber für die Formulierung seiner Forderungen gewesen ist.

Auch wird in der 1. Auflage der Methodenlehre der österreichische Psychiater und Nobelpreisträger Julius Wagner-Jauregg erwähnt, der der *Simultanmethode* als alternierendes Verfahren zur Gruppenbildung den Namen gegeben habe.

Bleuler und Wagner-Jauregg waren beide als Psychiater tätig. Martini scheint also wichtige Impulse für seine Ideen hinsichtlich der methodischen Durchführung therapeutischer Versuche aus dem Bereich der Psychiatrie erhalten zu haben. Im Teil „Darstellung der klinischen Forschung anhand exemplarischer Studien“ findet sich ein Beispiel, das auf ein geschärftes Bewusstsein für methodische Fragestellungen besonders in dem Bereich der

¹¹⁴ Martini 1932, S. 40.

Erforschung psychotroper Substanzen hinweist. Auch in der Methodenlehre finden sich Hinweise darauf und scheinen diesen Eindruck zu bestätigen.

Anschließend bietet es sich an, Veränderungen der Methodik in der zweiten. Auflage im Vergleich zur ersten der Methodenlehre zu überprüfen. Die zweite Auflage der „Methodenlehre“ wurde 1947 veröffentlicht, 15 Jahre nach der ersten, die 1932 also zu Anfang des untersuchten Zeitraums erschienen ist. Inhaltlich stimmen beide Auflagen im Wesentlichen überein.

Wie auch in der 2. schreibt Martini auch in der 1. Auflage schon, dass für die Ausschaltung von Mitursachen eine *unwissentliche* Versuchsanordnung, eine *isolierte* Betrachtung der Medikamentenwirkung und die Anwendung von *Ausschlusskriterien* notwendige Bedingungen seien.

Hinsichtlich der Voraussetzungen für die Beurteilung der Wirksamkeit von Medikamenten zeigen sich bemerkenswerte Veränderungen. Die Forderung, eine Vergleichsebene auf der Grundlage der Beobachtung des *zeitlichen Verlaufs* oder der *Vor- und Nachbeobachtungsperiode* zu bilden, findet sich in beiden Auflagen in annähernd unveränderter Weise wieder. Allerdings weicht die in der ersten Auflage beschriebene „*Simultanmethode nach Wagner-Jauregg*“ nicht nur von der Terminologie, sondern auch inhaltlich von dem Vorgehen ab, welches Martini in der zweiten empfiehlt. Seiner Ansicht nach könne auf der Grundlage der alternierenden Methode allein keine Homogenität innerhalb einer Gruppe gewährleistet werden und demnach sei eine „*Korrektur*“ erforderlich:

„Es ist also notwendig, sowohl die Fälle „alternierend“ zu verteilen und so der Zufallsverteilung zu folgen, als sie auch angesichts der drohenden Ungleichmäßigkeiten dieser Zufallsverteilung wegen doch noch zu ordnen. Praktisch läuft dies darauf hinaus, dass bei einer akuten Erkrankung – fast immer werden es Infektionskrankheiten sein – die augenblickliche zufällige Verteilung in Bezug auf Alter, Krankheitsstadium beim Einsatz der Therapie usw. dauernd in einer übersichtlichen Tabelle offenkundig gehalten wird, und dass diese Zufallsverteilung von Zeit zu Zeit bei den Krankenzugängen ausgeglichen wird, indem z. B. der Gruppe, die weniger Greise aufzuweisen hat, bei den nächsten Aufnahmen diese bevorzugt zugewiesen werden; entsprechend wäre auch in Bezug auf andere wichtige Merkmale zu verfahren.“¹¹⁵

¹¹⁵ Martini 1947 S. 16 f.

Diese Korrektur müsse aber nach Martinis Meinung mit Vorsicht angewandt werden und wird von ihm als „**Method der ausgleichenden Alternierung**“¹¹⁶ bezeichnet.

Gemessen an den heutigen Standards klinischer Forschung stellt die in der zweiten Auflage erwähnte „*Korrektur*“ im Vergleich zur ersten nicht unbedingt einen methodischen Fortschritt dar. Das liegt unter anderem auch daran, dass Martini für die von ihm vorgeschlagene „*Korrektur*“ keine eindeutigen und praktischen Regeln formuliert und somit die Gruppenbildung stärker der Gefahr der willkürlichen Zuordnung durch den Forscher aussetzt.

Hinsichtlich der mathematischen Auswertung präzisiert Martini seine Vorschläge vor allem, was den Stellenwert der Fehlerrechnung betrifft:

„Den Rang einer Wahrscheinlichkeit erhält die Differenz zweier relativer Häufigkeiten erst durch den Nachweis, dass sie sich auf einer „genügend großen Zahl“ von Beobachtungen aufbaut. Dieser Nachweis wird mit Hilfe der Fehlerrechnung geführt. Diese hat ganz allgemein die Aufgabe, zu prüfen, ob die mathematischen Grundlagen eines statistischen Resultates genügen, um das Resultat als zuverlässig erscheinen zu lassen.“¹¹⁷

In der Fehlerrechnung spielt der *mittlere Fehler* eine tragende Rolle:

„Eine direkte und sichere Auskunft über die Beweiskraft eines therapeutischen Ergebnisses, d. h. dafür, dass p_{diff} der Differenz der beiden relativen Häufigkeiten, der Rang einer Wahrscheinlichkeit zukommt, erhalten wir erst, wenn wir auch sie an ihrem mittleren Fehler messen.“¹¹⁸

Der Begriff des *mittleren Fehlers*, der anstelle des der *statistischen Streuung* tritt, wird nicht erst in der zweiten Auflage, sondern bereits 1934 in einem Aufsatz, der auszugsweise unten wiedergegeben wurde, von Martini verwendet. In diesem Artikel wurde auch die Bedeutung des mittleren Fehlers in ähnlicher Weise wie in der zweiten Auflage der Methodenlehre beschrieben.

Ansonsten zeigen sich beim Vergleich der Auflagen keine elementaren Unterschiede. So tauchen in beiden die Begriffe *arithmetisches Mittel*, *Zentralwert* und *häufigster Wert* auf. Lediglich die *spezifische Wahrscheinlichkeitszahl* wird in der zweiten Auflage durch den Begriff der Differenz der relativen Häufigkeiten - p_{diff} - ersetzt.

Insgesamt scheinen die Vorschläge speziell zur statistischen Auswertung in der zweiten gegenüber der ersten Auflage gereift zu sein. Das Vorgehen der *ausgleichenden*

¹¹⁶ Martini 1947, S. 16.

¹¹⁷ Martini 1947, S. 20.

¹¹⁸ Martini 1947, S. 21.

Alternierung scheint aus heutiger Sicht im Vergleich zu einer unkorrigierten Alternierung jedoch etwas problematischer.

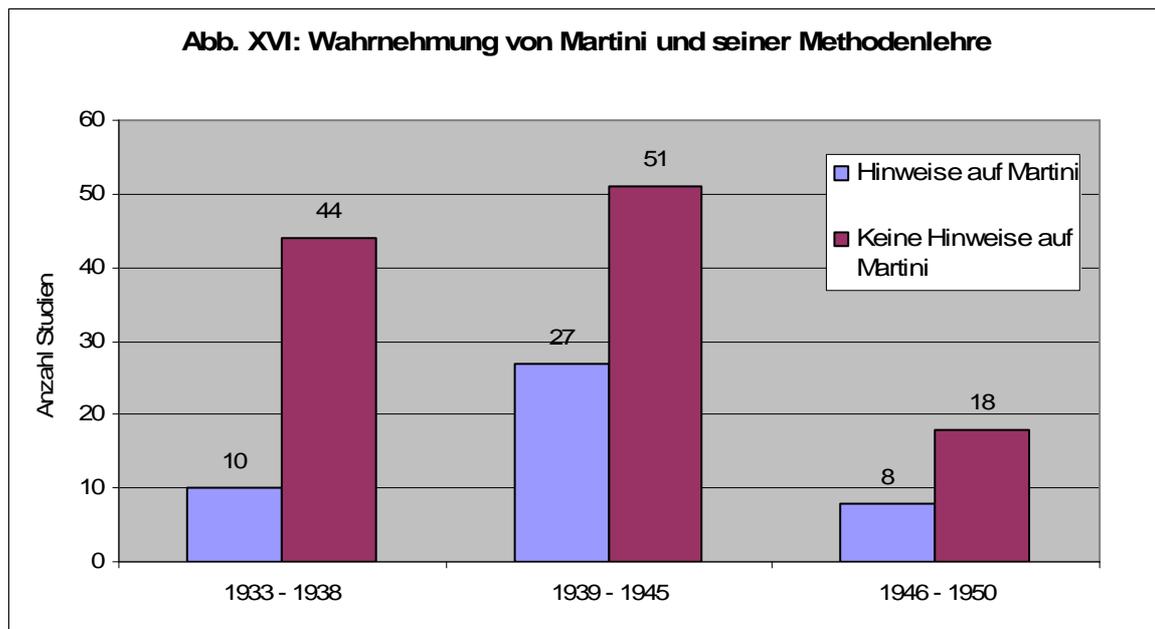
2. Wahrnehmung von Paul Martinis Ideen und der Methodenlehre

Im Folgenden wird untersucht werden, wie Martini und die Methodenlehre in den untersuchten Studien wahrgenommen wurden. Dafür wurden die Studien zunächst darauf untersucht, ob sich in den Texten oder in den Literaturangaben direkte Verweise auf Martini oder die Methodenlehre fanden. Die Studien, in denen keine direkte Bezugnahme auf Martini beschrieben wurde, wurden nach Hinweisen überprüft, die einen Bezug zu den in der Methodenlehre beschriebenen Forderungen erkennen lassen. So wurden Termini, die von Martini spezifisch verwendet wurden, als Indizien für die Wahrnehmung seiner Ideen betrachtet. Dazu zählen vor allem die Begriffe *Simultanmethode*, *alternierende* und *unwissentliche Versuchsanordnung*. Außerdem wurden Studien, in denen der Gebrauch des *mittleren Fehlers* beziehungsweise der *statistischen Streuung* nach den Empfehlungen von Martini beschrieben wurde, ebenfalls zu dieser Kategorie zugezählt.¹¹⁹

In 16 Studien, was einem Anteil von 10 % am Gesamtmaterial entspricht, wurde direkt auf Martini beziehungsweise die Methodenlehre eingegangen. In 29 (18 %) Veröffentlichungen wurden eine oder mehrere der oben erwähnten Maßnahmen beschrieben, die auf eine Umsetzung der Forderungen Martinis schließen lassen. In der überwiegenden Mehrzahl der Studien (72 %) konnten keine Hinweise diesbezüglich gefunden werden.

Als nächstes wurde überprüft, ob zwischen den zeitlichen Perioden Unterschiede in der Wahrnehmung Martinis erkennbar sind. In dieser Darstellung wurde auf die Differenzierung zwischen indirekten und direkten Verweisen der besseren Übersicht wegen verzichtet. Folgendes Diagramm zeigt die Verteilung:

¹¹⁹ Anmerkung: Dass Forderungen, wie Martini sie beschreibt, umgesetzt wurden, darf nicht als Beweis für einen direkten Zusammenhang zwischen dem Autor und Martini missverstanden werden.



Zu erkennen ist, dass in der ersten Periode, die Wahrnehmung von Martini im Vergleich zu den anderen zwei Perioden relativ betrachtet gering ist. Im Verhältnis zur Gesamtzahl wurden Martinis Ideen in der zweiten Periode am häufigsten implementiert. Ein etwas anderes Bild zeigt sich, betrachtet man nur die direkten Verweise auf Martini. Zwischen 1933 und 1938 wurde in zwei Studien ein direkter Bezug auf Martini beschrieben. In der zweiten Periode konnte in neun Studien ein direkter Verweis auf Martini beobachtet werden. In der 3. Periode nach Erscheinen der 2. Auflage wurde in fünf Studien direkt auf Martini eingegangen. Berücksichtigt man die Gesamtzahl der Studien, so zeigt sich dass in der 3. Periode in knapp 20 % aller Studien direkt auf Martini eingegangen wurde, während in der 1. Periode in etwa 4 % und in der 2. Periode in etwa 10 % der ausgewerteten Studien direkt auf Martini Bezug genommen wurde. Die deutliche relative Zunahme der Wahrnehmung von Martini beziehungsweise seiner Vorschläge deutet darauf hin, dass Martinis Forderungen auf wachsende Akzeptanz gestoßen sind.

Bei der Durchsicht des Materials wurde weiter untersucht, welche Vorschläge Martinis berücksichtigt wurden. In zehn von den insgesamt 16 Studien, in denen direkt auf Martini eingegangen wurde, bezogen sich die Autoren auf die Forderung Martinis, eine Vergleichsgrundlage bei der Studiengestaltung zu bilden. Fünf Autoren erwähnten, dass sie bei der statistischen Auswertung der Studienergebnisse die Empfehlungen Martinis berücksichtigt hätten.

Unter den 29 Studien, in denen sich Hinweise finden, dass Martinis Vorschläge berücksichtigt wurden, ohne aber dass dieser explizit erwähnt wurde, wurde in 17 Studien die Beschreibung des *alternierenden Verfahrens* oder anderer vergleichender Verfahren

beschrieben. In sieben Studien wurde eine *unwissentliche Versuchsanordnung* beschrieben, in drei Veröffentlichungen wurde die Berechnung von *statistischen Merkmalen*, wie sie von Martini gefordert wurde und in zwei Studien die *Ausschaltung von Mitursachen* durch sonstige Maßnahmen erwähnt. Insgesamt zeigt sich, dass insbesondere der Forderung Martinis nach einer Vergleichsgrundlage in Form des alternierenden Verfahrens in einem größeren Teil der Studien nachgekommen wurde. Weniger deutlich lässt sich das für seine Empfehlungen bezüglich der unwissentlichen Versuchsanordnung oder der statistischen Auswertung verfolgen.

3. Diskussion um Martinis Methodenlehre

Eine Veröffentlichung Martinis in der Klinischen Wochenschrift von 1934 führte zu einer kontroversen Diskussion über seine Empfehlungen hinsichtlich der statistischen Auswertung klinischer Ergebnisse. An dieser Stelle werden die wesentlichen Teile dieser Diskussion dargestellt. Martini schreibt in dem Artikel „Die Erfolgsbeurteilung therapeutischer Untersuchungen“:

„Die Forderungen, die ich in den letzten beiden Jahren für die Durchführung therapeutischer Untersuchungen angestellt habe, haben viel Zustimmung gefunden und haben auch Kritik erfahren. Die Kritik wird, soweit sie die Forderungen für zu streng oder undurchführbar hält, vor allem bei denen, die die Forderungen angehen, denen sie aber unangenehm sind, ein willkommenes Gehör finden als Zustimmung. Ich werde mich mit ihr an anderer Stelle auseinandersetzen. Der Teil der Kritik, der von statistischer Seite kommt, enthält wertvolle Anregungen, die besonders in wichtigen Forderungen der Anwendung der *Fehlerrechnung* ‚Methodenlehre therapeutischer Untersuchungen‘ ergänzen.“¹²⁰

In dem Teil „I. Beurteilung aus dem Krankheitsausgang.“ geht Martini auf die Berechnung und Bedeutung verschiedener statistischer Größen ein. So definiert er die „*spezifische Wahrscheinlichkeitszahl einer Therapie*“ als Kenngröße zur Bestimmung, ob eine spezifische Therapie einer unspezifischen überlegen ist:

„Die Anwendung der spezifischen Therapie hatte nur dann einen Wert, wenn die relative Häufigkeit der Heilung = die Wahrscheinlichkeit der Heilung unter der spezifischen Therapie größer war als bei unspezifischer Therapie, wenn also $w_T > w_s$ (Anmerkung: w_T : relative Häufigkeit der Heilung unter spezifischer Therapie; w_s : relative Häufigkeit der

¹²⁰ Martini 1934, S. 872.

Spontanheilung ohne spezifische Therapie). Die Größe der Differenz $w_T - w_s$ ist das Maß für den Grad der Überlegenheit der spezifischen Therapie über die unspezifische. Diese Differenz zweier relativer Häufigkeiten (der Heilung einerseits bei unspezifischer, andererseits bei spezifischer Therapie) habe ich die spezifische Wahrscheinlichkeitszahl einer Therapie genannt.“¹²¹

Um zu überprüfen, ob die „spezifische Wahrscheinlichkeitszahl“ eine genügend große Beweiskraft hat, schlägt Martini vor, diese mit dem „mittleren Fehler“ zu vergleichen: „Das Verhältnis von σ_{diff} zu der Wahrscheinlichkeitsdifferenz (= spezifische Wahrscheinlichkeitszahl) w_{diff} , also w_{diff}/σ_{diff} , ist maßgebend für die Beweiskraft der spezifischen Wahrscheinlichkeitszahl w_{diff} .⁷ w_{diff} muß σ_{diff} wiederum nicht nur um das 1 fache, sondern um wenn irgend möglich um das 3 fache übersteigen; erst dann kann von einem genügenden Beweis gesprochen werden.“¹²²

An anderer Stelle geht Martini auf die Kritik von einem ärztlichen Kollegen ein, der in einer früheren Stellungnahme Martinis Gebrauch des Begriffs *Wahrscheinlichkeit* kritisiert:

„HEIDENHAIN hat sich in einer ausführlichen Auseinandersetzung, ‚Klinische Erfahrung, Wahrscheinlichkeit und Fehlerrechnung‘, für die ihm alle, denen der Fortschritt der therapeutischen Untersuchung am Herzen liegt, dankbar sein müssen, gegen meine Anwendung des Begriffes der Wahrscheinlichkeit gewandt. Er geht dabei von einem Beispiel aus, das ich früher gebracht, und legt an diesem Beispiel dar, dass aus diesem Beispiel unmöglich eine ‚Wahrscheinlichkeit‘ errechnet werden könne.

Wahrscheinlichkeit müsse eine wenigstens annähernde Allgemeingültigkeit besitzen, die bei einer so kleinen Zahl von 30 bzw. 24 Fällen (und scil. einer kleinen Differenz der relativen Häufigkeiten) niemals erreicht werden könne.“¹²³

Martini kommt schließlich zu dem Schluss, dass die Kritik Heidenhains unberechtigt und seine eigene Verwendung der Begrifflichkeit gerechtfertigt sei:

„Tatsächlich ist die Übereinstimmung von relativer Häufigkeit und Wahrscheinlichkeit beim Vorliegen einer genügend großen Beobachtungsreihe eine vollkommene; beide sind dann identisch.“¹²⁴

In einem zweiten Artikel von Martini, der im 25. Heft in der Klinischen Wochenschrift 1934 erschienen ist und die Fortsetzung zum 1. Artikel darstellt, wird noch weiter auf das

¹²¹ Martini 1934, S. 873.

¹²² Martini 1934, S. 874.

¹²³ Martini 1934, S. 873.

¹²⁴ Martini 1934, S. 873.

mathematische Vorgehen bei Verlaufsbeurteilungen eingegangen. Hierbei schlägt Martini zur Erfolgsbeurteilung einer Therapie den Vergleich mit der Vorbeobachtungsperiode vor: „Das Urteil über den Wert oder Unwert des zu prüfenden Heilmittels kann auch hier gewonnen werden aus der *Differenz* von w_T und w_s (Anmerkung: bei w_T handelt es sich um das Verhältnis beobachteter Kriterien - z. B. Hämoglobin - zum Zeitraum während der Therapie, w_s beschreibt das Verhältnis der Kriterien zum Beobachtungszeitraum vor der Therapie). Man wird dann aber abhängig von dem jeweiligen und empirischen Maß des Kriteriums, so dass der Endwert z. B. verschieden wäre, je nachdem z. B. Hämoglobin bei der Behandlung und Beobachtung einer Anämie in absoluten Werten (in Gramm pro 100 ccm Blut) oder in Prozent gerechnet worden wäre. Die Abhängigkeit vom empirischen Maß hört auf, wenn statt der Differenz der *Quotient der beiden relativen Wahrscheinlichkeitswerte* w_s und w_T gebildet wird:

$$w_{\text{Quotient}} = w_T/w_s. \text{ „}^{125}$$

Im Folgenden erläutert Martini sein genaues Vorgehen, was aber keine wesentlichen Beiträge mehr zu der anschließenden Diskussion liefert.

Zunächst nimmt Siegfried Koller in dem Artikel „Die Erfolgsbeurteilung therapeutischer Untersuchungen“, der im August 1934 im 32. Heft erschienen ist, Stellung zu dem Aufsatz von Martini. Siegfried Koller war Mathematiker und leitete zu diesem Zeitpunkt die statistische Abteilung des Kerckhoff-Herzforschungszentrums in Bad Nauheim. Koller leitet seine Kritik wie folgt ein:

„Als Antwort auf die vielfache Kritik, die seine früheren Veröffentlichungen zur Frage der Beurteilung therapeutischer Maßnahmen erfahren haben, hat P. Martini in Nr. 24 und Nr. 25 dieser Wschr. (Anmerkung: Wochenschrift) eine neue Darstellung seiner statistischen Methodik gegeben. Manche Fehler der früheren Fassung sind jetzt richtig gestellt.

Insbesondere ist berücksichtigt, dass nicht Mittelwert- oder Quotientenbildung das Wesen eines kritischen Zahlenvergleichs bilden, sondern dass ein wirkliches Urteil nur durch die Berücksichtigung des natürlichen Schwankungsbereichs der Zahlen erfolgen kann.

So begrüßenswert auch diese Verlagerung des methodischen Schwerpunktes ist, kann man doch manches Grundsätzliche und viele Einzelheiten auch der neuen Darstellung nicht als richtig anerkennen. Gerade eine methodische Darstellung soll ja den, der sich nach ihren

¹²⁵ Martini 1934, S. 908.

Vorschriften richtet, zuverlässig beraten und vor groben Fehlern bewahren. Das ist aber auch bei diesem Aufsatz MARTINIS nicht der Fall.“¹²⁶

Im Detail sind es vier Punkte, die Koller an der Darstellung Martinis kritisiert. Zunächst sei der Gebrauch des *mittleren Fehlers* in Martinis Veröffentlichung aus Kollers Sicht unzulässig:

„Für kleinere Beobachtungsreihen dagegen ist die Anwendung des mittleren Fehlers nicht mehr zulässig. Vielmehr hat dann eine ausführliche Berechnung nach dem binomischen Satz zu erfolgen.“¹²⁷

Wie zuvor Heidenhain bemängelt auch Koller die Verwendung des in der Veröffentlichung Martinis gebrauchten Begriffs *Wahrscheinlichkeit*:

„Die Verwendung des Wahrscheinlichkeitsbegriffes bei MARTINI muss zu vielen Unklarheiten führen. Er sagt selbst, dass er diesen Begriff nicht nur im üblichen Sinne gebraucht, gibt aber keine Definition, was er sich dabei nun vorstellt.“¹²⁸

Weiter kritisiert Koller Martinis Vorgehen bei der Beurteilung einer therapeutischen Maßnahme für den Verlauf:

„Ein „statistischer Kunstfehler“ ist aber die Quotientenbildung, wenn der Kurvenverlauf betrachtet wird.“¹²⁹

Stattdessen fordert Koller:

„Es müssen also Differenzen geprüft werden, auch wenn diese Differenzen von der gewählten Maßeinheit abhängig sind.“¹³⁰

Zuletzt bemängelt Koller Folgendes:

„Noch eine weitere Angabe MARTINIS fordert scharfen Widerspruch heraus. Es ist dies die Bewertung einzelner besonders auffallender Beobachtungen. Wenn für solche ‚einzelne positive Resultate‘ zusammengenommen der dreifache mittlere Fehler überschritten ist, so sieht er für diese Fälle die Wirkung des Mittels für erwiesen an. Das widerspricht den Grundsätzen der Statistik, denn in jeder streuenden Beobachtungsreihe müssen einzelne Extreme auftreten.“¹³¹

Koller kommt letztlich zu folgendem Schluss:

¹²⁶ Koller 1934, S. 1154.

¹²⁷ Koller 1934, S. 1154.

¹²⁸ Koller 1934, S. 1154.

¹²⁹ Koller 1934, S. 1154.

¹³⁰ Koller 1934, S. 1154.

¹³¹ Koller 1934, S. 1155.

„Die statistischen Methoden MARTINIS zur therapeutischen Erfolgsrechnung sind auch in ihrer Neufassung zum großen Teil nicht richtig. Ihre Anwendung kann zu erheblichen Fehlschlüssen führen.“¹³²

Martini's Antwort auf die Kritik Kollers ist direkt im Anschluss an die Veröffentlichung abgedruckt. Er entgegnet der Kritik nach der Anwendung des *mittleren Fehlers*:

„KOLLER hat vor kurzem selbst erklärt, es lasse sich keine scharfe Grenze für die Anwendbarkeit des mittleren Fehlers und der Gauß-Verteilung geben.“¹³³

Der Vorwurf Kollers, er definiere den Begriff *Wahrscheinlichkeit* nicht, wird durch Martini entkräftet, indem er seine eigene Veröffentlichung zitiert:

„Unter Wahrscheinlichkeit ist in den folgenden, wie auch in meinen früheren Ausführungen nichts anderes verstanden als die relative Häufigkeit eines bestimmten Ausganges oder Verlaufes einer Erkrankung bei der Beobachtung einer genügend großen Zahl von Krankheitsfällen gleicher Art.“¹³⁴

Martini nennt zur Verteidigung der von Koller kritisierten Anwendung eines Quotienten zur Bestimmung des Einflusses eines Medikamentes auf den Krankheitsverlauf zwei Gründe:

„Ein Hauptgrund war eben der, dass eine Wahrscheinlichkeitsdifferenz, die aus einer Kurve gewonnen wird, tatsächlich zu keiner absoluten Wahrscheinlichkeitsdifferenz führt, und so von der mathematischen Wahrscheinlichkeitsrechnung nur mehr das Kleid borgt.(...) Der 2. Grund für meinen Vorschlag war die Abhängigkeit von empirischen Maßen bei der Errechnung von Wahrscheinlichkeitsdifferenzen; wenn man mit Quotienten arbeitet, wird man von ihnen unabhängig.“¹³⁵

Den letzten Einwand Kollers, er würde einzelne extrem Ergebnisse überbewerten, kommentiert Martini mit folgenden Worten:

„Kollers Ausführungen gehen hier gegen einen Feind, den er sich selbst konstruiert hat, und den er nur deshalb so ernsthaft bekämpfen kann, weil offenbar der medizinische Unsinn, den er mir zuschreibt, dem Mathematiker nicht in seiner ganzen Größe klar geworden ist.“¹³⁶

Koller findet in den einleitenden Worten durchaus Lob für Martini's Bemühen, einige Ungenauigkeiten, die in der 1. Auflage der Methodenlehre enthalten sind, zu verbessern.

¹³² Koller 1934, S. 1155.

¹³³ Martini 1934, S. 1155.

¹³⁴ Martini 1934, S. 1155.

¹³⁵ Martini 1934, S. 1155 f.

¹³⁶ Martini 1934, S. 1156.

Dennoch erscheint die Kritik Kollers an Martini durchaus schwerwiegend zu sein. Gerade die abschließenden Worte Kollers, lassen doch erkennen, dass er zu einem insgesamt schlechten Urteil gelangt, was die Empfehlungen Martinis zur statistischen Auswertung betrifft. Wenig überraschend ist daher die Vehemenz, mit der Martini die eigenen Vorschläge verteidigt und seinerseits Koller für dessen aus seiner Sicht vorschnelle und ungerechtfertigte Kritik angreift. Während in dem Meinungs austausch zwischen Koller und Martini vor allem inhaltliche Fragen dominieren, fällt die Diskussion zwischen Heidenhain und Martini durch einen sehr persönlichen und zum Teil unsachlichen Stil auf.

Als Antwort auf die von Martini geäußerte Kritik an Heidenhain veröffentlicht selbiger eine Richtigstellung, die 1934 in der Klinischen Wochenschrift im 38. Heft erschienen ist. Heidenhain selbst war als Chirurg in Worms tätig und hat seinerseits einige Veröffentlichungen zur statistischen Auswertung von klinischen Ergebnissen herausgegeben. Er leitet seinen Aufsatz wie folgt ein:

„Um den Leser ins Bild zu bringen, eine kurze Vorbemerkung: MARTINI schrieb eine Methodenlehre der therapeutischen Untersuchung. Er stellt allgemein mit weiten Gesichtspunkten und mathematisch-rechnerisch den Vergleich von Behandlungsmethoden dar. Mathematisch ging er fehl. (...) Darum muss ich hier einige allgemeine Gesichtspunkte herausheben, um mich gegen MARTINIS Darstellung an oben genannter Stelle zu verteidigen.“¹³⁷

Heidenhains Kritik an Martinis Aufsatz „Die Erfolgsbeurteilung therapeutischer Untersuchungen“ besteht im Wesentlichen aus drei Punkten. Zunächst greift Heidenhain Martinis Definition von **Wahrscheinlichkeit** an:

„Denn Wahrscheinlichkeit errechnet der Mathematiker aus vielen Zehntausenden von ‚Fällen‘, unterste Grenze für die Praxis des Lebens 2000. Über derartig große Zahlen verfügen wir in unseren klinischen Erfahrungen fast nie. Für gewöhnlich sind es höchstens einige hunderte. Die Bezeichnung W (Wahrscheinlichkeit, d. i. wahrscheinliche Richtigkeit) für Relativzahlen (z/n und $100 z/n$) ist also grundfalsch.“¹³⁸

Weiter sind die mathematischen Operationen, die Martini vorschlägt, Ziel der Kritik:

„Hiervon abgesehen ist MARTINIS Formel (3) $w_{\text{diff}} = w_T - w_S$ rein rechnerisch falsch. Bei Vergleich der Zahlen spezifisch und nicht spezifisch behandelter Kranken handelt es sich um eine *Wertbestimmung durch Zahlen*. *Wertvergleiche* kann nie durch Subtraktion,

¹³⁷ Heidenhain 1934, S. 1360.

¹³⁸ Heidenhain 1934, S. 1360.

sondern nur durch Division stattfinden. Vergleicht man die Werte 5 und 2, so erhält man nicht $5 - 2 = 3$, sondern $5 : 2 = 2,5 : 1$.¹³⁹

Auch spricht sich Heidenhain gegen die von Martini empfohlene Art der Bildung von **Kontrollgruppen** aus:

„Man erhält oft durch jahrelange Versuche mit Methoden oder Dosierungen überwiegend negative Ergebnisse, bis man auf den richtigen Weg kommt. Klinische Versuche pflegt man – ist man mit den Ergebnissen der üblichen Methode unzufrieden – nicht derart anzustellen, dass man, wie MARTINI (fiktive Beispiel 2 und 3, S. 875) (Anmerkung: In diesen Beispielen beschreibt Martini die Bildung von Kontrollgruppen durch alternierende Verfahren), abwechselnd spezifisch und unspezifisch behandelt. Wir Chirurgen wenigstens tun dies nicht.“¹⁴⁰

Abschließend kommt Heidenhain zu dem Ergebnis:

„I. Der ‚mathematische Entwicklungsgang, um zu einem therapeutischen Urteil zu gelangen‘, wie MARTINI sich ausdrückt, in der Methodenlehre beschrieb und auch jetzt noch verteidigt, ist nach elementaren rechnerischen Begriffen falsch, hat mit Mathematik nichts zu tun.“¹⁴¹

In Martinis Erwiderung, die direkt im Anschluss an die Richtigstellung von Heidenhain abgedruckt ist, leitet er seine Verteidigung gegenüber der Kritik Heidenhains wie folgt ein: „Dass aber ein Arzt die Bezeichnung von ‚relativer Häufigkeit‘ als ‚Wahrscheinlichkeit‘ wiederholt und mit Leidenschaft für grundfalsch erklärt, im wesentlichen deshalb, weil wir in der Medizin nicht wie der Mathematiker über wenigstens einige 1000 sondern höchstens über einige 100 von ‚Fällen‘ verfügen, das ist deprimierend, weil es die Gefahr mit sich bringt, dass im Streit um Worte die Sache Schaden leide. Ich erkläre heute zum 3. Male: ‚Unter Wahrscheinlichkeit ist in den folgenden, wie auch in meinen früheren Ausführungen nichts anderes verstanden als die relative Häufigkeit eines bestimmten Ausganges oder Verlaufes einer Erkrankung bei der Beobachtung einer genügend großen Zahl von Krankheitsfällen.‘“¹⁴²

Dem Vorwurf, er verwende unzulässige mathematische Operationen, entgegnet Martini: „Wenn ich gegenüber KOLLER mit gewissen Vorbehalten den Quotienten zweier relativer Häufigkeiten in Schutz genommen habe, so muss ich ohne jede Einschränkung

¹³⁹ Heidenhain 1934, S. 1360.

¹⁴⁰ Heidenhain 1934, S. 1361.

¹⁴¹ Heidenhain 1934, S. 1361.

¹⁴² Martini 1934, S. 1361.

HEIDENHAINS Angriffe gegen die Differenz zweier relativer Häufigkeiten (bzw. zweier Wahrscheinlichkeiten) als völlig irrig zurückweisen und KOLLER durchaus zustimmen, dass es am nächsten liegt, die *Differenz* zu betrachten und zu prüfen, ob sie sicher oder nur durch den Zufall der kleinen Zahl von Null verschieden ist'.¹⁴³

Die Kritik Heidenhains an der von Martini vorgeschlagenen Art zur Gruppenbildung, bezieht Martini vor allem auf die Frage nach der Vorbeobachtung.

„Was schließlich HEIDENHAIN über ‚Kontrollreihen‘ und ‚negative Ergebnisse‘ ausführt, das richtet sich gegen die Notwendigkeit der Vorbeobachtung. Seine Feststellung, dass die Chirurgen ‚dies bisher nicht getan haben‘ dürfte er selbst kaum als Beweis dafür halten, dass es deshalb so in Ordnung gewesen wäre.“¹⁴⁴

Es ist nicht sicher, ob Martini an dieser Stelle die Kritik Heidenhains missverstanden hat. Denn Heidenhain scheint seine Kritik vor allem gegen das alternierende Verfahren zur Gruppenbildung zu richten.

Martini greift seinerseits einige Arbeiten Heidenhains an, die seiner Ansicht nach methodisch mangelhaft seien. In dieser Auseinandersetzung zwischen Martini und Heidenhain scheinen nicht nur rein inhaltliche Differenzen eine Rolle zu spielen.

Wie bereits in dem Kapitel Material und Methoden erläutert, ist die Checkliste, die zur Evaluation der Studien diente, stark an die Forderungen Martinis angelehnt. Daher sollte gezeigt werden, welche Probleme Martini in der klinischen Forschung sah und mit welchen Vorschlägen er die Probleme lösen wollte. Die Resultate zur Wahrnehmung Martinis deuten auf ein wachsendes Interesse an dessen Ideen hin. Insbesondere Martinis Forderung nach einer Vergleichsgrundlage für die Beurteilung klinischer Resultate fand unter den klinischen Forschern Anklang. Ein zusätzliches Indiz für das Interesse an methodischen Fragen zur Studiendurchführung kommt schon allein durch die Länge des Artikels „Die Erfolgsbeurteilung therapeutischer Untersuchungen“ zum Ausdruck. Der Verlag stellte dem Artikel einen relativ großen Raum von zehn Seiten und der anschließenden Diskussion einen Platz von fünf Seiten zur Verfügung.

Die Diskussion zwischen Martini, Koller und Heidenhain zeigt, dass die Methodenlehre Martinis interdisziplinär kritisch diskutiert wurde. Zustimmung erhält Martini vor allem für sein Bemühen, methodische Grundlagen für klinische Forscher zu formulieren. Die Vorschläge Martinis zur statistischen Auswertung werden jedoch sowohl von Koller als

¹⁴³ Martini 1934, S. 1361.

¹⁴⁷ Martini 1934, S. 1361.

auch von Heidenhain als mangelhaft betrachtet, wobei Koller und Heidenhain in ihrer Kritik nicht etwa übereinstimmen. Gerade diese unterschiedliche Auffassung zur statistischen Auswertung lässt eine gewisse Unsicherheit unter den klinischen Forschern im Umgang mit statistischen Größen erkennen und macht darüber hinaus deutlich, dass Martinis Versuch, eine einheitliche methodische Basis für die klinische Forschung zu schaffen, einen notwendigen Schritt darstellte, um diese Probleme zu überwinden. Heidenhains Angriff auf den Vorschlag Martinis, vergleichende Untersuchungen zur Erfolgsbeurteilung durchzuführen, zeigt außerdem, dass auch die methodische Umsetzung klinischer Forschung keinesfalls als unumstritten angesehen werden kann.

5. Untersuchung externer Einflüsse auf die klinische Forschung

Dieses Kapitel gliedert sich in zwei Teile. Im ersten Teil wird untersucht, wie Forschung aus dem Ausland von den deutschen Forschern wahrgenommen wurde. Außerdem soll überprüft werden, ob sich Veränderungen der Wahrnehmung ausländischer Literatur zeigen. Im zweiten Teil dieses Kapitels wird auf die Rolle der klinischen Studien aus dem nicht zivilen Umfeld und ihre Besonderheiten eingegangen.

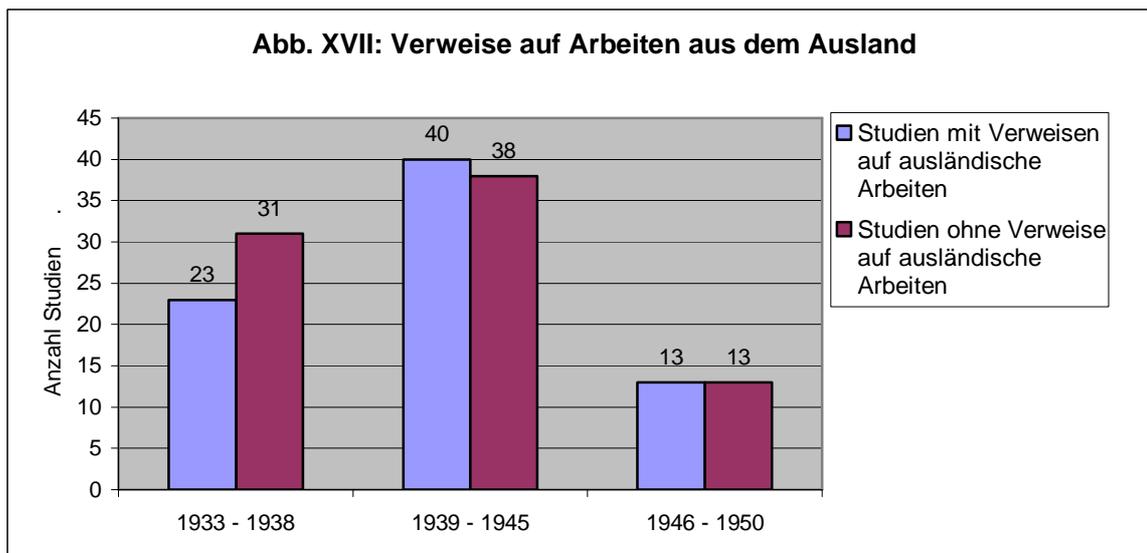
1. Einfluss der Forschung aus dem Ausland

Die Checkliste von Stoll wurde um das Evaluationsmerkmal „Rezeption aus dem Ausland“ erweitert, so dass die Wahrnehmung ausländischer Literatur in dem zugrunde liegenden Material entsprechend verfolgt werden konnte. Bei der Evaluation des genannten Kriteriums wurde weiter differenziert, welchem Land die Autoren der zitierten Literatur entstammten. Aufgrund von ungenauen Literaturangaben und der Übersicht wegen wurde allerdings nur zwischen angelsächsischen Autoren und Autoren aus anderen Ländern unterschieden. Wie auch bei den anderen untersuchten Merkmalen konnten die Studien mit Angaben aus dem Ausland den verschiedenen Perioden zugeordnet und im zeitlichen Verlauf überprüft werden.

In insgesamt 76 Studien wurde Literatur aus dem Ausland zitiert. Das entspricht einem Anteil von etwa 48 %. In den übrigen 82 Studien (52 %) fanden sich keine Hinweise, dass Autoren aus dem Ausland zur Kenntnis genommen wurden. 53 (69 %) von den 76 Studien, in denen ausländische Verfasser zitiert wurden, beziehen sich auf Arbeiten aus dem angloamerikanischen Sprachraum.

Unter den 23 Studien (30 %), die Autoren aus anderen Ländern zitieren, finden sich sechs Studien, in denen auf Autoren aus der Schweiz eingegangen wurde und fünf Studien, die französische Forscher zitieren. In neun Studien ließen sich keine Erkenntnisse über die genaue Herkunft der Autoren gewinnen. In drei Studien, die von ausländischen Forschern durchgeführt wurden, wurden Ergebnisse deutscher Arbeiten zitiert.

Die Studien, nach dem oben beschriebenen Merkmal untersucht, verteilen sich wie folgt auf die verschiedenen Perioden:



Betrachtet man die ersten beiden Perioden, so zeigt sich, dass der Anteil der Publikationen, in denen von ausländischer Forschung berichtet wird, von 43 % auf 51 % anwächst. In der Nachkriegszeit wird in 50 % der Veröffentlichungen auf Forschung aus dem Ausland eingegangen. Bedenkt man, dass in der Kriegszeit der Zugang zu Journalen aus dem Ausland gegenüber der Vor- oder der Nachkriegszeit erschwert war, scheint diese Entwicklung bemerkenswert zu sein.

Die Auswertung des Materials widerspricht der Annahme, dass sich die klinische Forschung in Deutschland zur Kriegszeit zunehmend introvertiert und von äußeren Einflüssen isoliert hatte. Tatsächlich deuten die Ergebnisse sogar eher auf das Gegenteil hin. Es scheint das Interesse an der Forschung der Kollegen aus dem Ausland gegenüber der Vorkriegszeit zu wachsen. Gleichzeitig mit der Wahrnehmung inhaltlicher Fragen wird auch die methodische Durchführung klinischer Studien aufgefallen sein. Auch wenn man vermuten kann, dass das Interesse vor allem inhaltlichen Problemen und weniger methodischen gegolten hat, so ist es doch wahrscheinlich, dass die in den Artikeln aus dem Ausland beschriebene Methodik deutschen Forschern bekannt gewesen sein dürfte, die diese Literatur zur Kenntnis genommen haben.

Unter den 158 ausgewerteten Publikationen befanden sich auch Artikel aus dem Ausland auf teilweise sehr hohem methodischem Niveau. Diese Artikel sind insofern besonders interessant, weil sie einer relativ großen Anzahl von Medizinerinnen und klinischen Forschern in Deutschland zugänglich waren. Daher soll folgend ein Artikel aus Dänemark zusammenfassend dargestellt und die angewandte Methodik erläutert werden.

Im Jahr 1938 erscheint in der Klinischen Wochenschrift ein Artikel von Poul Bahnsen, einem dänischen Forscher, unter dem Titel „Studien über die Wirkung von β -

Phenylisopropylaminsulfat (Mecodrin) auf normale Menschen“. Bahnsen ging in seinen Versuchen der Frage nach, inwieweit sich die stimulierende Wirkung des Symathomimetikums Mecodrin auf die Erhöhung der geistigen Leistungsfähigkeit auswirkte. Dazu muss angemerkt werden, dass es sich um keine klinisch therapeutische Studie an Patienten handelt, sondern dass die Versuchspersonen gesund waren.

So beschreibt Bahnsen bei der Darstellung der Untersuchungsmethode:

„Etwa 300 – 400 Männer und Frauen im Alter von 17 – 63 Jahren, im wesentlichen Kontorpersonal verschiedener öffentlicher und privater Institutionen in Kopenhagen, haben sich für die Versuche zur Verfügung gestellt. Der Arbeitsalltag war folgender: ehe die Versuche begannen, wurden die Versuchspersonen in Bezug auf Blutdruck und Herzfunktion objektiv untersucht und um frühere Krankheiten befragt. Nur normale Personen wurden angenommen.“¹⁴⁵

Zur Versuchsdurchführung schreibt Bahnsen:

„Vor den eigentlichen Versuchen wurde mit Hilfe dieser Fragebogen eine Gruppe von Personen untersucht, die überhaupt kein Präparat bekamen, um zu sehen, wie das subjektive Befinden an einem zufälligen Tag unter normalen Verhältnissen beurteilt wurde. Diese Gruppe (Normalgruppe) gab uns nicht nur einen Einblick darin, wie die Fragebogen von den Versuchspersonen ausgefüllt wurden, sondern vor allem ein großes Material zum Vergleich mit den Angaben der Versuchspersonen. Die Versuche wurden teils mit *Mecodrin* und teils mit Neutraltabletten gleicher Form und Größe und gleichen Geschmacks (0,1 mg Chininsulfat pro Tablette) wie die Mecodrintabletten durchgeführt. Die männlichen Versuchspersonen bekamen 4 Tabletten (20 mg), die weiblichen 2 (10 mg). Der größte Teil der Versuchspersonen wusste zwar, dass sie entweder Mecodrin, oder Kontrolltabletten bekommen konnten, aber nicht, was ihnen verabreicht wurde. Eine kleinere Gruppe wurde in dem Glauben gelassen, dass sie Mecodrin bekämen, obgleich man einigen von ihnen Kontrolltabletten gab. Bei einem Teil der Versuchspersonen wurden die Versuche mehrmals mit willkürlichen Wechsel de Mecodrin- und Kontrolltage wiederholt.“¹⁴⁶

An der Darstellung dieser Versuchsanordnung ist insbesondere die Art der Durchführung der **Placebokontrolle** hervorzuheben. Bahnsen beschreibt explizit und ausführlich die Anwendung von Scheinmitteln zum einen an einer größeren Gruppe und zum anderen in abwechselnder Folge in der Verumgruppe. Da sich das Scheinmittel und das Mecodrin sowohl in Geschmack wie scheinbar auch in Form ähneln, konnten die Probanden keine

¹⁴⁵ Bahnsen 1938, S. 1075.

¹⁴⁶ Bahnsen 1938, S. 1075.

sensorischen Unterschiede feststellen, was er ebenfalls explizit im Versuchsaufbau erwähnt. Durch Täuschung der Gruppe, welche er ausschließlich mit Scheinmitteln behandelt, ihnen aber suggeriert, dass sie mit Mecodrin behandelt werden würden, bildet Bahnsen eine Placebogruppe. Am Ende verfügt Bahnsen über drei verschiedene Gruppen.¹⁴⁷ Die Bildung der „*Normalgruppe*“, die keinerlei Beeinflussung erfahren hat, verfolgt Bahnsen mit dem Ziel, eine Vergleichsgrundlage zur Bewertung der Beeinflussung durch Mecodrin zu gewinnen. In der „*Kontrollgruppe*“ wurden die Probanden zusammengefasst, denen durch das Scheinmittel suggeriert wurde, sie würden mit Mecodrin behandelt werden. Die dritte Gruppe ist die Mecodrin-Gruppe selbst.

Dass Bahnsen einen für die Zeit vergleichsweise sehr innovativen Ansatz des Einsatzes von Scheinmitteln zeigt, deutet auf ein sehr sensibles Bewusstsein für methodische Fragestellungen hin. Zum Ausdruck kommt dieses Bewusstsein nicht nur bei der Beschreibung der Versuchsanordnung, sondern auch bei der Darstellung der Ergebnisse: „Wie vorher erwähnt, suggerierten wir darum einer kleinen Gruppe bei der Verabreichung von Kontrolltabletten, dass sie Mecodrin bekämen, das leicht stimulierend wirke, die Leistungsfähigkeit und das allgemeine Wohlbefinden erhöhe und die Müdigkeit vermindere. Die Gruppe, die nach dieser Instruktion wirklich Mecodrin erhielt, zeigte in der Beantwortung der Fragen keine wesentlichen Abweichungen von der großen Mecodringruppe. Ebenso wenig wichen die Wirkungen der Kontrolltabletten bei dieser Gruppe in Bezug auf Anzahl oder Intensität wesentlich von der großen Kontrollgruppe ab. (...) Hieraus kann man also mit größtmöglicher Sicherheit schließen, dass die beobachteten Symptome nicht auf die Kenntnis der zu erwartenden Wirkung des Präparates seitens der Versuchspersonen zurückzuführen ist.“¹⁴⁸

Bahnsen macht hier deutlich, dass er sich über die zusätzliche Möglichkeit der Ergebnisauswertung durch die Bildung der Placebokontrollgruppe sehr wohl im Klaren ist nämlich, dass er zwischen der physiologischen oder der tatsächlichen Wirkung des Mecodrins und der psychologischen Wirkung bei Erwartung der Verabreichung von Mecodrin differenzieren kann.

Die beschriebene Studie ist unter methodischen Aspekten als vergleichsweise hochwertig anzusehen. Eine *placebokontrollierte* Studie dieser Art setzt eine gründliche Auseinandersetzung mit methodischen Fragestellungen voraus. Bahnsen begründet zudem sein Vorgehen sehr detailliert, so dass dem Leser die methodischen Probleme bei der

¹⁴⁷ Vgl. Bahnsen 1938, S. 1075.

¹⁴⁸ Bahnsen 1938, S. 1076.

Versuchsgestaltung verdeutlicht werden. Unter diesem Aspekt erfüllt der Artikel auch eine Art fortbildende Funktion auf dem Gebiet der Methodik.

Außerdem stellt diese Publikation ebenso ein Beleg dafür dar, dass es in Dänemark klinische Forscher gegeben hat, die für die damaligen Verhältnisse ein zukunftsweisendes Studiendesign angewendet haben.

Zusammenfassend kann festgestellt werden, dass die Wahrnehmung von Studien aus dem Ausland selbst in Kriegszeiten gewachsen ist und dass man daher davon ausgehen kann, dass methodische Neuentwicklungen aus dem Ausland nicht unbemerkt von deutschen Forschern geblieben sein dürften.

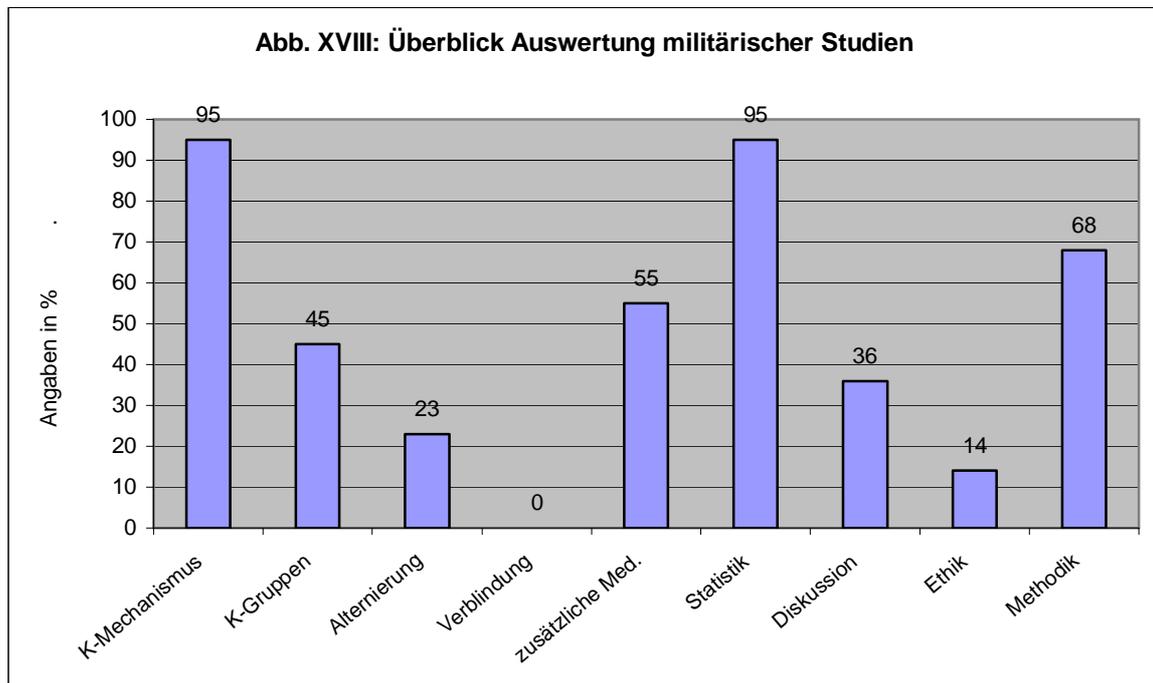
2. Einfluss aus nicht-zivilen Einrichtungen

Auffällig bei der Untersuchung der Verteilung der Anzahl der Studien auf die Perioden ist, wie bereits festgestellt wurde, die deutliche Zunahme von klinisch-therapeutischen Studien in der Kriegszeit. Für diese Zunahme scheinen unter anderem auch Studien aus dem militärischen Umfeld verantwortlich zu sein. Eine genauere Auseinandersetzung mit dieser Subgruppe der klinischen Forschung soll den Einfluss des nicht-zivilen Kontexts auf die Gestaltung der klinischen Studien untersuchen.

Es finden sich in dem untersuchten Material insgesamt 22 klinische Studien, die in militärischen Einrichtungen, in Arbeits- oder Gefangenenlagern durchgeführt wurden. Dies entspricht einem Anteil von ungefähr 14 % am Gesamtmaterial. Von den 22 Studien verteilen sich drei auf die Vorkriegszeit, 17 auf die Kriegs- und zwei auf die Nachkriegszeit. Bei den zwei Studien in der Nachkriegszeit handelt es sich allerdings um Arbeiten, die zu Zeiten des Krieges entstanden sind, aber erst nach Kriegsende veröffentlicht wurden. Daher können diese Arbeiten entsprechend ihrem Entstehungszeitpunkt auch der Kriegsperiode zugerechnet werden.

Setzt man die Anzahl der Studien mit militärischem Hintergrund in Relation zur Anzahl der Gesamtstudien in der jeweiligen Periode, so liegt der Anteil der Studien mit militärischem Hintergrund in der Vorkriegszeit noch bei knapp 6 % und vervierfacht sich auf 22 % in der Kriegsperiode. In der Nachkriegsperiode befinden sich, wenn man die zwei erwähnten Arbeiten herausrechnet, keine militärischen Studien mehr. Elf (50 %) von diesen 22 Studien wurden als Outcomesstudie identifiziert und zehn (45 %) als experimentelle Studien. Unter den 22 Studien befand sich außerdem eine Fallstudie.

Evaluert man die Studien hinsichtlich verschiedener Studienmerkmale, so zeigt sich folgendes Ergebnis:



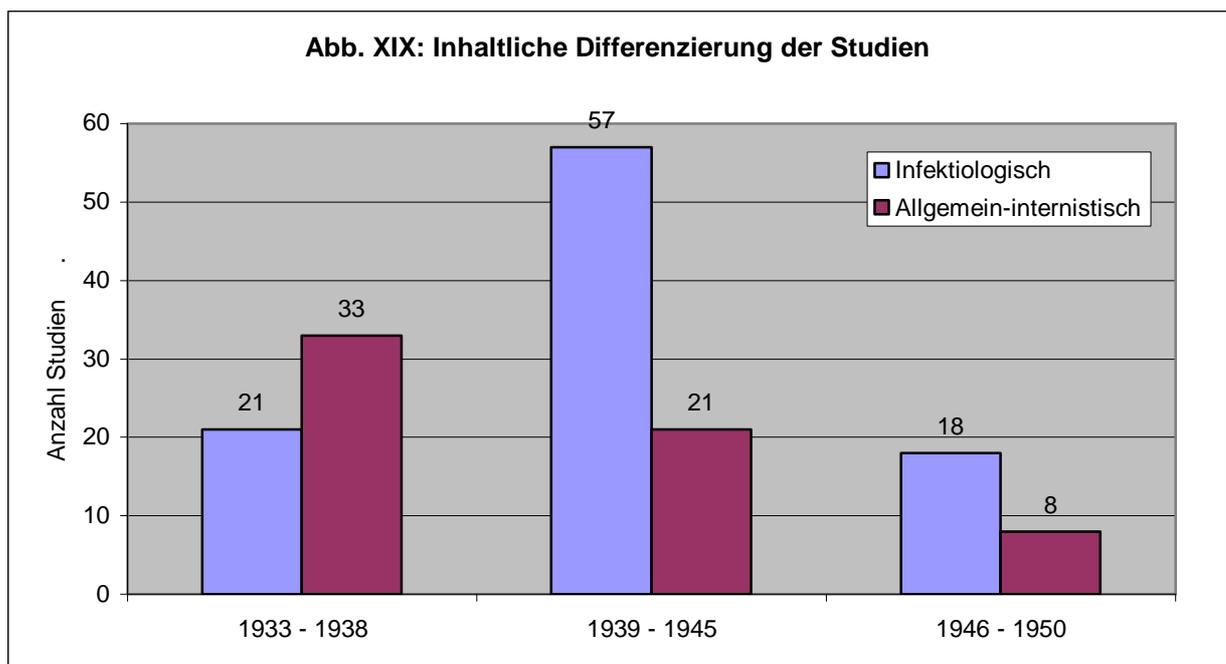
Die Anwendung von Kontrollmechanismen wurde mit einer Ausnahme in allen Studien erwähnt. Kontrollgruppen wurden in fast der Hälfte der Studien beschrieben. Ein alternierendes Verfahren bei der Gruppenbildung wurde in knapp 25 % aller Studien erwähnt. Eine Verblindung der Versuche konnte in keiner der Studien nachgewiesen werden. Die Erfassung von Störgrößen durch Erwähnen zusätzlich verabreichter Medikamente wurde in etwas mehr als der Hälfte der Lazarettstudien nachgewiesen. Eine statistische Auswertung erfolgte in nahezu allen Publikationen. In drei Studien (14 %) wurde eine komplexere statistische Methode durchgeführt. Eine kritische Auseinandersetzung kann in knapp 40 % der Studien festgestellt werden, eine methodische in 68 %. In drei Studien (14 %) wird auf ethische Aspekte der Forschung Bezug genommen.

Vergleicht man die ausgewerteten Merkmale innerhalb dieser Gruppe mit denen der Gesamtstudien, so zeigen sich besonders für die Merkmale der Alternierung, der Erfassung von Störgrößen und der statistischen Auswertung Unterschiede.

Ein alternierendes Verfahren ließ sich im Gesamtmaterial von 158 Studien bei 22 erkennen, was etwa 14 % entspricht. Der Anteil der Studien aus einem militärischen Umfeld, in denen ein alternierendes Verfahren bei der Gruppenbildung angewandt wurde, ist für knapp 25 % aller Studien mit alternierendem Verfahren verantwortlich. Eine statistische Auswertung wird in 70 % aller Studien beschrieben. Demgegenüber wird in nahezu allen

militärischen Studien eine statistische Auswertung durchgeführt. Die Erfassung von Störgrößen durch Beschreibung möglicher Interaktionen mit anderen zusätzlich verabreichten Medikamenten wird in dem Gesamtmaterial in ca. 40 % beschrieben. Innerhalb der Studien mit militärischem Charakter kann dieses Kriterium in über der Hälfte der Studien beobachtet werden. Für die anderen Kriterien lassen sich keine wesentlichen Unterschiede feststellen.

Als indirekte Folge des Krieges ist die Zunahme von Infektionskrankheiten in den Kriegsjahren zu betrachten. Das lässt sich auch an dem Material verfolgen. Unterteilt man das Material in Studien, die dem Bereich der *Infektiologie* zuzuordnen sind und solchen, die sich mit *allgemein internistischen* oder anderen nicht infektiologischen Fragestellungen befassen und überprüft, wie sich die so differenzierten Arbeiten auf die Perioden verteilen, so zeigt sich folgendes Bild:



In dem Diagramm lässt sich erkennen, dass sich etwa 40 % der Gesamtstudienzahl der Vorkriegszeit mit der Erforschung zur Behandlung infektiologischer Erkrankungen auseinandersetzen. In dem Kriegszeitraum wächst dieser Anteil auf 73 % an und bleibt in der Nachkriegszeit etwa auf diesem Niveau. Innerhalb des militärischen Umfelds befassen sich 18 von 22 (ca. 80 %) der Studien mit infektiologischen Fragestellungen. Insgesamt beträgt der Anteil infektiologischer Studien am Gesamtmaterial etwa 60 %.

Bereits in der Vorkriegsperiode befasst sich ein großer Teil mit der Erforschung von Behandlungsmöglichkeiten von Infektionserkrankungen. Gut wirksame Breitspektrum

Antibiotika wie zum Beispiel Penicillin standen im klinischen Alltag noch nicht zur Verfügung. Folglich waren Infektionskrankheiten schwer therapierbar. Insofern erklärt sich, dass das Interesse zur Bekämpfung infektiologischer Erkrankungen im Fokus vieler klinischer Forscher stand.¹⁴⁹

In der Kriegsperiode wächst der Anteil infektiologischer Studien von etwa 40 % auf über 70 % an. In diesem Anstieg spiegelt sich die zunehmende Verschlechterung der hygienischen Situation in der Zivilbevölkerung als auch in den militärischen Einrichtungen wider.

Ein möglicher Grund, der für die steigende Anzahl von Studien im Bereich der Infektiologie diskutiert werden kann, ist das Interesse der militärischen Führung, über wirksame Antibiotika zu verfügen, um die Wehrtauglichkeit der Soldaten zu gewährleisten und wiederherzustellen.¹⁵⁰ Der große Anteil infektiologischer Studien im nicht-zivilen Umfeld weist jedenfalls daraufhin, dass der antibiotischen Forschung eine übergeordnete Rolle zukam und vermutlich als kriegswichtig eingeschätzt wurde.

Aus den dargestellten Ergebnissen lassen sich einige Besonderheiten der klinischen Forschung in einem nicht zivilen Umfeld verfolgen. Der Anteil der militärischen Forschung an den Gesamtstudien beträgt in der Kriegsperiode knapp ¼. Die militärische Forschung hat in dieser Zeit für die klinische Forschung eine nicht zu unterschätzende zahlenmäßige Bedeutung. Inhaltlich verschiebt sich in diesem Zeitraum der Fokus von einer allgemein internistischen zu einer überwiegend infektiologischen Forschung. Das lässt sich am ehesten durch kriegsbedingte Umstände erklären. Versucht man die Forschung auf der Grundlage der untersuchten Merkmale qualitativ zu beurteilen, so lässt sich feststellen, dass die Forschung in dem militärischen Umfeld eine vergleichsweise gehobene Qualität aufweist. Das gilt insbesondere für die Zuweisung zu Kontrollgruppen und die Anwendung statistischer Methoden. In folgenden Beispielen sollen exemplarisch einige erwähnenswerte methodische Voraussetzungen der klinischen Forschung im militärischen Kontext erläutert werden.

In dem Artikel „Ulcusleiden und vegetatives Nervensystem“, der 1949 in der Zeitschrift *Medizinisch Klinik* erschienen ist, veröffentlicht K. Nitsch eine Studie, die im Jahr 1944 durchgeführt wurde. Nitsch schreibt:

¹⁴⁹ Vgl. Loudon 2002.

¹⁵⁰ Vgl. Kanovitch 1997, S. 94 f.

„Ich habe mich nun bemüht, in einer Einheit von etwa 850 Magenkranken den statistisch gesicherten Beweis zu erbringen, daß es möglich ist, durch Dämpfung des vegetativen Nervensystems Ulcusrezidive zu verhindern.

Alle 324 Männer waren in völlig gleiche Verhältnisse im Rahmen einer im Küstenschutz eingesetzten Einheit gezwungen. Ihre Kost war gleichartig und kann mit dem Schlagwort einer „Schonkost“ ohne strenge Einschränkungen bezeichnet werden. Gelegentliche „Diätfehler“ durch Erwerb ungeeigneter Lebensmittel werden sich annähernd gleichmäßig auf alle später zu schildernden Gruppen verteilt haben, ebenso die sehr verschiedenartige seelische Belastung dieser letzten Kriegsjahre. (...) Diese 324, weitgehend unter gleichen äußeren Bedingungen lebenden Männer wurden in drei Gruppen geteilt, und zwar nach ihrer alphabetischen Aufstellung der erste in Gruppe I, der zweite in Gruppe II, der dritte in Gruppe III usw. Diese Aufteilung war z. B. so gleichmäßig, dass die größte Differenz im Durchschnittsalter, die zwischen Gruppe II und III bestand, nur 9/12 Jahre betrug. (...) Auch die Durchschnittsgewichte wie auch die Größe waren so gleichartig, dass es sich erübrigt, Zahlen zu bringen.“¹⁵¹

Nitsch beschreibt einen alternierenden Versuchsaufbau. Der Unterschied zu den meisten anderen Studien, in denen alternierend zugewiesen wurde, besteht darin, dass Nitsch nicht etwa zeitlich, sondern alphabetisch alterniert. Er betont, dass die äußeren Bedingungen für alle Teilnehmer ungefähr gleich gewesen seien. Die drei gebildeten Gruppen seien außerdem, was das Alter und die körperliche Konstitution betrifft, etwa gleich.

Nitsch fährt fort:

„Die während der Beobachtungszeit ausfallenden 27 (wegen andersartiger Krankheiten, Verletzungen usw.) verteilten sich annähernd gleichmäßig (...) Gruppe I blieb ohne irgendwelche Behandlung, Gruppe II erhielt täglich 3 mal 1 Tablette Bellergal und 3 mal 1 Luminatablette, zeitweise aus Mangel an Bellergal 3 mal 2 Luminatabletten, und zwar immer 3 Wochen lang, dann 1 Woche Pause. Bei Gruppe III wurde dieselbe Medikation nur dann gegeben, wenn entweder ein Gewichtsabfall (mehr als 300 g pro Woche) oder eine deutliche Zunahme der vegetativen Dystonie festzustellen war. Die Beobachtung erfolgte vom 1.3.1944 bis zum 1.12.1944.“¹⁵²

Der Autor beschreibt die Anwendung von Ausschlusskriterien in Form von „*andersartigen Verletzungen*“. Die Gruppe I diente ihm als Kontrollgruppe insbesondere gegenüber Gruppe II und teilweise gegenüber Gruppe III. Außerdem grenzt er präzise den Beobachtungszeitraum ein. Nitsch erwähnt eher beiläufig, dass der Mangel an

¹⁵¹Nitsch 1949, S. 503.

¹⁵²Nitsch 1949, S. 503.

Medikamenten dazu geführt hat, dass die Behandlung der Gruppe II nicht in der geplanten Art durchgeführt werden konnte.

Nitsch beschreibt außerdem so etwas wie eine doppelte Verblindung der Versuchsanordnung:

„Um jeden Subjektivismus des Arztes möglichst auszuschließen, war diesem nicht bekannt, welcher der drei Gruppen der einzelne angehörte. (...) Den Männern der Gruppen II und III war nicht bekannt, aus welchem Grunde sie ihre Tabletten schlucken mussten. (...) Wenn ein Patient wegen Magenbeschwerden näher untersucht wurde, wusste ich nicht, welcher Gruppe er angehörte. Sobald ein Ulcus festgestellt wurde, schied der Betreffende aus der weiteren Gruppenüberwachung aus.“¹⁵³

Nitsch erwähnt, dass den Versuchsteilnehmern unbekannt war, aus welchem Grund sie Tabletten bekamen. Er betont weiter, dass auch er selbst nicht über die Gruppenzugehörigkeit und damit über die Behandlung der einzelnen Probanden Bescheid wusste.

Dieser Artikel ist stellvertretend für eine Reihe von klinischen Studien aus einem nicht-zivilen Rahmen zu betrachten. Zum einen ist er wegen der angewandten Methodik und zum anderen wegen Besonderheiten im Umgang mit militärischem Personal erwähnenswert. In der von Nitsch beschriebenen Methodik werden die *Alternierung*, die Anwendung von *Ausschlusskriterien* und die Erwähnung einer *doppelt blinden* Versuchsanordnung erwähnt. Für diesen Zeitraum ist die gewählte Methodik als fortschrittlich zu betrachten. Insbesondere die explizite Erwähnung einer doppelt blinden Studienanordnung verdient Beachtung.

Im Vergleich zu klinischen Studien im zivilen Rahmen gibt es unterschiedliche Voraussetzungen, die auf die angewandte Methodik Einfluss haben. Einige dieser für militärische Studien typischen Bedingungen, die sie von zivilen Studien abgrenzen, lassen sich aus der Publikation von Nitsch ableiten.

Zunächst ist die *Rekrutierung* von Versuchsteilnehmern in zivilen und militärischen Bereichen verschieden. Im militärischen Bereich rekrutieren sich die Versuchsteilnehmer aus einem relativ homogenen Kollektiv. Im Unterschied zu einem gemischt geschlechtlichen Kollektiv in Krankenhäusern kann man davon ausgehen, dass in militärischen Einrichtungen fast ausschließlich männliche Versuchspersonen an therapeutischen Studien teilnehmen. Bezüglich des Alters der Teilnehmer lässt sich vermuten, dass es bei militärischen Personen homogener ist als bei Personen, die in zivilen

¹⁵³ Nitsch 1949, S 503.

Einrichtungen behandelt werden und die für das Merkmal Alter eine größere Streuung aufweisen. Wegen der heterogenen Altersverteilung unter den Studienteilnehmern in zivilen Einrichtungen ergibt sich vermutlich auch eine größere Häufigkeit vom Auftreten von Begleiterkrankungen unter Zivilpersonen. Es ist daher davon auszugehen, dass sich der Forderung nach einem möglichst homogenen Patientenkollektiv in militärischen Einrichtungen leichter nachkommen ließ als in zivilen. Um in Studien mit Zivilpersonen eine möglichst homogene Gruppe zu bilden, ist die kompromisslose Anwendung von Rekrutierungs- und Ausschlusskriterien entscheidend, da das Kollektiv, aus dem sich die Versuchsgruppen zusammensetzen, an sich wesentlich heterogener ist als in militärischen Einrichtungen.

Für die Beurteilung des *zeitlichen Verlaufs* einer Erkrankung ergeben sich ebenfalls unterschiedliche Voraussetzungen. Im zivilen Rahmen beginnt die Beobachtung einer Krankheit und ihrer Behandlung in der Regel mit der Einweisung in eine medizinische Einrichtung und endet mit der Entlassung. Es ist ein gewisser Aufwand notwendig, um die Beobachtung zeitlich auszuweiten. Aufgrund der Zugehörigkeit zu einer militärischen Einheit lässt sich in nicht zivilen Studien diese Beobachtung leichter auf die Zeit vor und nach einer Behandlung ausweiten.

Lebensbedingungen, wie zum Beispiel Ernährung und Ähnliches, lassen sich für Zivilpersonen meist nur im Rahmen eines stationären Aufenthalts kontrollieren. Für Angehörige des Militärs ist eine längere und präzisere Erfassung der *Lebensbedingungen* eher denkbar.

Die unterschiedlichen Voraussetzungen zwischen zivilen und militärischen Studien lassen einige nicht unwesentliche, methodische Vorteile zugunsten von militärischen Studien erkennen. Das gilt insbesondere für das homogene Gesamtkollektiv als auch für die zeitliche Erfassung und Kontrolle.

Von Studien, bei denen Angehörige des Militärs teilnahmen, müssen die Studien abgegrenzt werden, die in Gefangenenlagern durchgeführt wurden. In dem untersuchten Material befinden sich zwei Studien, die an Lagerinsassen durchgeführt wurden. An dem Beispiel des Artikels „Chemo- und Serotherapieversuche an Flecktyphuskranken“, der in der Klinischen Wochenschrift 1942 von Rudolf Wohlrab dem Leiter der Fleckfieberabteilung am staatlichen Institut für Hygiene zu Warschau veröffentlicht wurde, sollen Besonderheiten klinischer Versuche unter Gefängnisinsassen aufgezeigt werden. Ziel ist es, vor allem auf ethische Probleme der Forschung in diesem Kontext hinzuweisen.

Wohlrab schreibt:

„Wir haben 51 Flecktyphuskranke mit *Prontosil rubrum* bzw. *Rubiazol* behandelt (Rychard) und fanden eine Verkürzung der Fieberzeit um 1,2 Tage (14,92 Tage Fieber) gegenüber 61 unbehandelten, alternierenden Kontrollkranken (16,08 Tage Fieber). Leider ist das bei den großen Temperaturschwankungen, die bei den einzelnen Kranken festzustellen sind, kein sicheres Ergebnis. Der klinische Eindruck war entsprechend unsicher oder negativ. Mit beiden Präparaten wurde bei einer Gefängnisepidemie prophylaktisch behandelt (Rychard). Von den 19 täglich mit 0,5 – 1,5 g *Rubiazol* behandelten Gefängnisinsassen erkrankten 8 mit typischem Krankheitsverlauf (Fieberdauer 15,6 Tage), von 20 alternierenden Kontrollen nur 7 (Fieberdauer 14,5 Tage).“¹⁵⁴

In den folgenden Zeilen geht Wohlrab dann auf die Ergebnisse seiner Studie ein, ohne weitere Details zur Studiendurchführung mitzuteilen.

Für einen Forscher aus dem zivilen klinisch-therapeutischen Bereich ist die Einhaltung ethischer Normen ein verpflichtendes Prinzip. Zweifelhaft ist, ob sich Forschende aus nicht zivilen Einrichtungen ebenfalls diesen Prinzipien verpflichtet fühlten. Im Besonderen gilt das für Gefängnis- und Lagereinrichtungen. Vor dem Hintergrund dieser Überlegung, muss kritisch gefragt werden, ob sich der in einem Gefangenenlager Forschende aus einer ethischen Motivation heraus der Heilung der Versuchspersonen verpflichtet fühlte oder aus einem rein wissenschaftlichen Interesse heraus handelte. Die Verwendung des Begriffs „*Gefängnisinsasse*“, der den Begriff „Patient“ ersetzt, deutet auf eine Werteverchiebung im Umgang mit den Versuchsteilnehmern hin.

Hinsichtlich der tatsächlichen Motivation von Wohlrab lassen sich schwerlich Aussagen treffen. Dennoch ist die angewandte Begrifflichkeit auffällig.

Wohlrab erwähnt beiläufig ein alternierendes Verfahren bei der Bildung der Gruppen. Wie dieses alternierende Verfahren im Einzelnen ausgesehen hat und welche Kriterien bei der Bildung der Gruppenbildung angewandt wurden, lässt sich nicht aus der kurzen Darstellung heraus verfolgen. Dennoch lässt sich vermuten, dass bei der Auswahl von Kriterien in Studien in einem klinischen Umfeld Unterschiede zu der Wahl von Kriterien bei Studien in einem Gefängnislager angewandt wurden. Die Besonderheiten hinsichtlich der ethischen Haltung von Lagerärzten werfen vor allem auch die Frage auf, ob sich der Forscher in einer vergleichenden Studie überhaupt zu einem Behandlungserfolg der Kontrollgruppe verpflichtet sah. Vor diesem Hintergrund ist es denkbar, dass in einigen Studien die Kontrollgruppe unbehandelt geblieben ist und billigend ein Schaden

¹⁵⁴ Wohlrab 1942, S. 455.

oder eventuell sogar das Versterben des „Versuchsteilnehmers“ in Kauf genommen worden ist.¹⁵⁵

Wohlrab schreibt weiter:

„Die ersten 33 Behandelten zeigten gegenüber ihren zur gleichen Zeit in den gleichen Krankensälen liegenden 31 Kontrollen eine Fieberverkürzung um 2,62 Tag.“¹⁵⁶

Die Erwähnung, dass auch die Kontrollgruppe in Krankensälen untergebracht wurde, weist in dieser Studie zumindest daraufhin, dass auch die Kontrollgruppe einer Behandlung zugeführt wurde, obwohl Wohlrab keine explizite Behandlung der Kontrollgruppe beschreibt.

Zweifellos haben die Folgen des zweiten Weltkrieges einen erheblichen Einfluss auf die klinische Forschung gehabt. Das zeigt sich zunächst an der Zunahme von Studien aus einem nicht-zivilen Umfeld während des Krieges. Auch die geringe Anzahl von klinischen Studien nach Kriegsende, die zum Teil durch das Nichterscheinen von medizinischen Zeitschriften bedingt ist, ist als unmittelbare Folge des Krieges zu verstehen. Auch inhaltlich lässt sich ein Einfluss beobachten. So wurde die Erforschung von **antibiotischen Medikamenten** forciert, was mit Sicherheit im Interesse der Militärführung gewesen sein dürfte. In einigen Studien wurde außerdem erwähnt, dass ein Mangel an Medikamenten herrschte. Diese Beobachtungen stellen zweifellos wichtige Indizien dafür dar, dass der Krieg Einfluss auf die klinische Forschung gehabt hat.

Aus methodischer Sicht lässt sich zeigen, dass militärische Studien zum Teil **ideale Voraussetzungen** für klinische Versuche aufweisen. Ein weiterer Aspekt wird in dem Artikel von Wohlrab deutlich. Klinische Studien aus dem nicht-zivilen Bereich weisen besondere **ethische Grundlagen** auf. Das zeigt sich in allererster Linie an der Begrifflichkeit und der sich daraus ableitenden Haltung des Forschers gegenüber den Versuchsteilnehmern.

¹⁵⁵ Vgl. dazu auch Schmiedebach 1987, S. 198 ff.: Schmiedebach beschreibt, dass eine zusätzliche Funktion des nationalsozialistisch beeinflussten Gesundheitssystems in dem Erhalt und der Wiederherstellung der „Volksgesundheit“ bestand. Diese Beobachtung weist auf eine problematische ethische Umgebung klinischer Forschung hin.

¹⁵⁶ Wohlrab 1942, S. 455

6. Diskussion

Eine zentrale Frage der vorliegenden Arbeit lautet: Lässt sich eine gerichtete methodische Entwicklung der klinisch therapeutischen Forschung in Deutschland zwischen 1933 und 1950 beobachten? Des Weiteren soll, wie in der Einleitung beschrieben, die Beurteilung der klinischen Forschung durch Martini in dem Vorwort zur 2. Auflage der Methodenlehre geprüft werden. Darin äußert er sich enttäuscht über den geringen Effekt der Methodenlehre auf die klinische Forschung.¹⁵⁷

Um sich einer Antwort auf diese Fragen anzunähern, müssen verschiedene Aspekte berücksichtigt werden.

Zunächst ist kritisch zu hinterfragen, ob mit der angewandten Methodik eine Annäherung an diese Fragestellung überhaupt möglich ist. Bereits im Abschnitt „Material und Methoden“ wurde auf verschiedene Schwierigkeiten bei der Durchführung der Arbeit hingewiesen.

Die Auswahl der Zeitschriften, die sich an der von Stoll¹⁵⁸ orientiert, stellt ein erstes Problem dar. Mit der Deutschen Medizinischen Wochenschrift, der Münchner Medizinischen Wochenschrift und der klinischen Wochenschrift wurden mit Sicherheit drei der bedeutendsten medizinischen Fachzeitschriften in Deutschland berücksichtigt. Ergänzt wurde die Selektion durch zwei kleinere Zeitschriften das „Archiv für klinische Medizin“, die „Verhandlungen der deutschen Gesellschaft für innere Medizin“ und die Medizinische Zeitschrift, die als Ersatz der DMW, MMW und der klinischen Wochenschrift in der Zeit zwischen Oktober 1944 und Februar 1945 erschienen ist. Allerdings wurde darauf verzichtet andere Blätter wie zum Beispiel die „Medizinische Klinik“ oder die „Medizinische Welt“ in die Auswertung mit einzubeziehen. Insofern wurden nicht alle klinisch therapeutischen Studien aus dem genannten Zeitraum durch die Checkliste erfasst und es kann kein Anspruch auf Vollständigkeit erhoben werden. Das mit Sicherheit größte Problem der Arbeit bestand darin, dass einige Merkmale sich einer objektiven Beurteilung stärker entziehen als andere. Auf die Schwierigkeit der Abgrenzung zwischen wissenschaftlicher Studie und Erfahrungsbericht mit eher narrativem Charakter wurde bereits im Teil Material und Methoden aufmerksam gemacht. Auch andere zu bewertende Eigenschaften der Studien ließen sich nur schwer objektiv beurteilen. Fällt es zum Beispiel relativ leicht, in Studien festzustellen, ob zusätzliche

¹⁵⁷ Martini 1947 im Vorwort.

¹⁵⁸ Vgl. Stoll 2005 S. 25 f.

Medikamente verabreicht wurden, ist es schwieriger, Kriterien zu definieren, die objektiv nachvollziehbar erkennen lassen sollen, ob eine kritische oder methodische Auseinandersetzung innerhalb der Studie erfolgt ist. Somit sind einige Aspekte der Checkliste stärker von der individuellen Beurteilung des Untersuchers abhängig als andere. Zudem existiert für die Literaturrecherche und die Studiaauswertung keine externe Kontrolle. Es wurden zwar in Zusammenarbeit mit Raspe und Roelcke einige Studien exemplarisch ausgewertet, um generelle Verständnisfragen bezüglich der Analyse zu klären. Dennoch stellt dieses Vorgehen keinen Ersatz für eine umfassende Interrater-Reliabilitätsprüfung von Recherche und Analyse dar.

Weiterhin muss darauf hingewiesen werden, dass die zum Teil lückenhafte Darstellung der Veröffentlichungen eine präzise Bewertung aller Aspekte der Checkliste nicht zulässt. Teilweise wurde diesem Umstand dadurch Rechnung getragen, dass zwischen expliziten und impliziten Hinweisen im Text unterschieden wurde. Dennoch können auch mit Anwendung dieser Differenzierung nicht alle Merkmale der Checkliste eindeutig beurteilt werden.

Trotz dieser Einschränkungen lassen sich wertvolle Informationen aus dem Material gewinnen, um eine Annäherung an die Frage nach methodischen Veränderungen in der klinischen Forschung zu ermöglichen.

Die Vielzahl an untersuchten Merkmalen bietet die Möglichkeit, Veränderungen auf verschiedenen Ebenen der klinischen Forschung abzubilden. Die Ähnlichkeit der Methodik zwischen dieser Arbeit und der von Stoll lässt außerdem einen Vergleich beider Arbeiten zu und gestattet es, zeitliche Verläufe zu verfolgen, was insbesondere wegen der zu prüfenden Bedeutung von Paul Martinis Forderungen an die therapeutische Forschung wichtig ist. Ergänzend finden sich in der Arbeit Auszüge aus verschiedenen Veröffentlichungen, um einen lebhaften Eindruck von der Umsetzung klinischer Forschung zu vermitteln.

Für den Zeitraum zwischen 1933 und 1950 wurden in der Arbeit 158 Veröffentlichungen evaluiert. Entsprechend errechnet sich ein Wert von etwa neun Studien pro Jahr. Bezieht man die Studienanzahl auf die verschiedenen Perioden, so zeigt sich, dass zwischen 1933 und 1938 54 Studien ausgewertet wurden entsprechend neun Studien pro Jahr. In der Kriegszeit wurden 78 Studien evaluiert also jährlich elf und in der Nachkriegszeit bis 1950 sind insgesamt 26 Veröffentlichungen also fünf pro Jahr evaluiert worden.

Berücksichtigt man, dass die Jahre 1944 und 1945 aufgrund der massiven Zerstörungen

durch den fortgeschrittenen Krieg gesondert zu betrachten sind, errechnet sich für die Jahre von 1939 bis 1943 eine durchschnittliche Publikationsanzahl von sogar etwa 14 Studien pro Jahr. Der starke Rückgang der Studienanzahl in der Nachkriegszeit lässt sich am ehesten als Folge des Krieges erklären. Die Zunahme der durchschnittlichen Anzahl klinischer Studien pro Jahr, die noch deutlicher ausfällt, wenn man die Jahre 1944 und 1945 aus der Betrachtung ausschließt, gegenüber der Vorkriegsperiode kann aber paradoxerweise ebenfalls als Folge des Krieges verstanden werden. Dieser scheinbare Widerspruch lässt sich am ehesten dadurch erklären, dass in den Kriegsjahren etwa 24 % der insgesamt ausgewerteten Studien in einem militärischen Rahmen durchgeführt worden sind, während es in der Vorkriegszeit lediglich 6 % der Studien mit militärischem Hintergrund waren. Hinter der starken Zunahme nicht-ziviler Studien verbarg sich in erster Linie das Interesse nach der Erforschung von wirksamen Antibiotika. So befassten sich 82 % der in diesem Zeitraum erschienenen militärischen Studien mit infektiologischen Fragestellungen. Auch die zivile Forschung beteiligte sich zunehmend an der Erforschung des Potentials neu entwickelter Antibiotika.¹⁵⁹ Spätestens ab 1944, als sich der Kriegsschauplatz immer stärker den Grenzen des Reichsgebiets genähert hat, ist von einer gravierenden Zerstörung der Infrastruktur auszugehen, was schließlich zu einer dramatischen Abnahme klinischer Studien geführt hat. Einen drastischen Beleg dafür liefern die Ausführungen von Adolf Bingel, als er in einer Veröffentlichung schreibt: „Die letzte Reihe hatte ich zahlenmäßig ausgewertet, als durch Feindeinwirkung eine Anzahl Krankengeschichten vernichtet wurde. Auf Einzelheiten konnte daher nicht mehr eingegangen werden.“¹⁶⁰

Vor diesem Hintergrund kann der Krieg gerade zu Beginn als Triebfeder für die therapeutische Forschung verstanden werden, da er zu einem erheblichen Teil zur Zunahme klinischer Studien geführt hat. Erst in den letzten Kriegsjahren kehrt sich dieser Effekt um und führt zu einer Abnahme der Anzahl klinischer Studien bis weit nach 1945.

Der Vergleich stellt ohne Zweifel einen wichtigen Bestandteil des klinisch therapeutischen Versuchs im modernen wissenschaftlichen Sinn dar und drängt sich daher als zentrales Kriterium zur Überprüfung der Frage nach einem methodischen Fortschritt auf.

In dem Material lässt sich in etwa 90 % der Studien ein Kontrollmechanismus, der einen Vergleich möglich macht, in verschiedenen Formen feststellen. Vergleicht man die

¹⁵⁹ Vgl. Loudon 2002: Auf die enorme Bedeutung der Erforschung von Antibiotika nach 1930 macht Loudon am Beispiel der Sulfonamide aufmerksam.

¹⁶⁰ Bingel 1949 S. 103.

verschiedenen Perioden miteinander, lässt sich keine eindeutige Entwicklung zugunsten der Anwendung von Kontrollmechanismen beobachten. Sind es in der Vorkriegsperiode fünf Studien (9 %) ohne Kontrollmechanismus, wird in neun Studien (12 %) der Kriegsperiode keine Kontrolle beschrieben. In der Nachkriegszeit fällt der Anteil von Studien ohne Kontrollmechanismus mit 8 % am geringsten aus.

Für die Anwendung von Kontrollgruppen in klinischen Studien zeigt sich ebenfalls kein einheitliches Bild. Beträgt der Anteil von experimentellen Studien in der Vorkriegsperiode knapp 41 %, wächst er in der Kriegsperiode auf 46 % an. In der Nachkriegsperiode schrumpft der Anteil auf nur noch 35 %. Bei der Gruppenbildung wurde in 23 von 67 kontrollierten Studien ein alternierendes Verfahren eingesetzt, was einem Anteil von 34 % entspricht.¹⁶¹ Im Gebrauch dieses Verfahrens zeigt sich, dass während der Vorkriegsperiode eine alternierende Gruppenbildung in vier von 22 kontrollierten Studien (18 %), in der Kriegsperiode in 14 von 36 (39 %) und in der Nachkriegsperiode in fünf von neun kontrollierten Studien (56 %) beschrieben wurde. Zumindest an dem Beispiel der Anwendung der alternierenden Methode mag man einen gewissen Fortschritt erkennen. Bereits 1835 wurde einer der ersten placebokontrollierten therapeutischen Versuche im deutschsprachigen Raum in den Nürnberger-Kochsalzversuchen beschrieben.¹⁶²

Vor diesem Hintergrund erscheint es überraschend, dass sich in dem gesamten Material von 158 Publikationen nur vier placebokontrollierte Studien finden, obwohl das Instrument der Placebokontrolle bereits seit mehr als 100 Jahren bekannt war. Ordnet man die vier Publikationen den Perioden zu, stellt sich heraus, dass drei von den vier Studien in der Vorkriegszeit und eine in der Kriegszeit erschienen sind, während sich in der Nachkriegszeit kein placebokontrollierter Versuch auffinden lässt. Inhaltlich befassen sich alle vier Studien mit Fragestellungen, die der Erforschung von psychotropen Mitteln gewidmet sind, was auf die besondere Rolle der Neurowissenschaften als Impulsgeber für methodische Fortschritte in klinisch-therapeutischen Studien hinweist.¹⁶³

In den Nürnberger-Kochsalzversuchen wird nicht nur eine Placebokontrolle erwähnt, sondern darüber hinaus wurde auch die Verblindung sowohl von den Studienteilnehmern

¹⁶¹ Vgl. Stoll 2005, S 106: Stoll bezeichnet das alternierende Verfahren als „Quasirandomisierung“.

¹⁶² Vgl. Löhner 1835.

¹⁶³ Vgl. Kaptchuk 1998, S. 408 ff.: Kaptchuk beschreibt, dass verblindete Prüfungen bereits Anfang des 19. Jahrhunderts in Versuchen eingesetzt wurden, die dem Wirkungsnachweis von homöopathischen Mitteln dienten. Im ausgehenden 19. Jahrhundert wurden diese Versuchselemente auch in Prüfungen mit psychotropen Mitteln eingesetzt. Erst für Anfang des 20. Jahrhunderts wird der Einsatz von Verblindungen auch in Versuchen mit allgemein internistischen Fragestellungen beschrieben.

als auch von den Untersuchern beschrieben. Die doppelblinde Versuchsanordnung war also ebenfalls bereits seit spätestens 1835 im deutschsprachigen Raum bekannt.¹⁶⁴

Zu dem Merkmal der verblindeten Versuchsanordnung zeigt sich eine ähnlich geringe Umsetzung in dem Material. Nur sieben von 158 Studien beschreiben eine verblindete Versuchsanordnung, wobei lediglich eine einfache Verblindung in den Studien beschrieben wird, nicht aber ein doppelblinder Versuchsaufbau. Von diesen insgesamt sieben verblindeten Versuchen sind drei vor dem Krieg und vier im Krieg erschienen entsprechend 6 % in der Vor- und 5 % in der Kriegsperiode. In der Nachkriegsperiode findet sich lediglich eine Studie, in der eine Verblindung erwähnt wird, was sich aber auf einen Versuch von 1918 bezieht und daher nicht in der Auswertung berücksichtigt wurde.¹⁶⁵

Die Anwendung statistischer Verfahren lässt sich ebenfalls als ein geeignetes Merkmal betrachten, um methodische Fortschritte in der klinischen Forschung zu messen. Bezüglich der Ergebnisauswertung zeigt sich, dass in ca. 60 % des Materials überwiegend einfache mathematische Operationen wie die Berechnung von Mittelwerten eingesetzt wurden. 28 % der Autoren verzichteten auf mathematische Darstellungen und Analysen ihrer Ergebnisse. In etwa 12 % kamen komplexere statistische Verfahren zum Einsatz. Dabei darf man den Begriff einer komplexen statistischen Berechnung nicht an den heutigen Ansprüchen messen, sondern muss beachten, dass zum Beispiel die Anwendung des mittleren Fehlers bei der Auswertung bereits als fortschrittlich zu verstehen ist. Für das Merkmal der statistischen Auswertung lässt sich ebenfalls keine eindeutige Entwicklung beobachten. So wächst der Anteil der Studien mit komplexeren statistischen Operationen zwar von der Vor- bis zur Kriegsperiode von 6 % auf 16 % an, stagniert in der Nachkriegszeit aber mit etwa 15 % auf dem Level der Kriegsperiode.

Überprüft man abschließend das Bewusstsein für methodische Fragestellungen als einen wichtigen Indikator für einen Fortschritt, zeigt sich, dass der Anteil der Publikationen, in denen methodische Fragen diskutiert wurden, von 52 % in der Vorkriegsperiode auf 68 % in der Kriegszeit anwächst, dann aber in der Nachkriegsperiode wieder auf 58 % sinkt.

Legt man das heutige Verständnis der Durchführung klinischer Versuche als Bewertungsmaßstab zugrunde, ist ein Fortschritt der Methodik in dem Beobachtungszeitraum nicht zu erkennen.

Speziell zwischen der Vor- und der Kriegsperiode deuten sich fortschrittliche Veränderungen sowohl in quantitativer wie auch in qualitativer Hinsicht an. Diese

¹⁶⁴ Vgl. Löhner 1835

¹⁶⁵ Vgl. Bingel 1949 S. 101 - 103

positiven Veränderungen lassen sich vor allem für den zunehmenden Anteil von Experimenten und der wachsenden Anwendung mathematischer Operationen bei der Ergebnisauswertung beobachten. Auch scheint zwischen den beiden Perioden das Bewusstsein für methodische Probleme insgesamt zuzunehmen.

Dennoch lässt sich keine nachhaltige positive Entwicklung bemerken. Zwischen der Kriegs- und Nachkriegsperiode erweist sich der Großteil der positiven Veränderungen für die meisten Merkmale als kurzfristig. So nimmt der Anteil experimenteller Studien an dem Gesamtmaterial in der Nachkriegszeit ab, die Anwendung statistischer Methoden stagniert auf dem Niveau der Kriegszeit, methodische Fragestellungen wurden deutlich seltener diskutiert. Andere innovative Merkmale, wie zum Beispiel der Einsatz von Placebokontrollen oder die Verblindung von Studienteilnehmern, wurden in der Nachkriegszeit überhaupt nicht mehr umgesetzt, was umso bemerkenswerter ist, als dass bereits in der Vorkriegsperiode mehrere beispielhafte Versuche für die Umsetzung solcher Verfahren in größeren Zeitschriften veröffentlicht wurden und somit deren methodische Durchführung und die Vorteile dieser prinzipiell bekannt waren. Lediglich für den Einsatz des alternierenden Verfahrens bei kontrollierten Versuchen lässt sich ein gewisser bleibender Fortschritt erkennen.

Dennoch können diese punktuellen Verbesserungen nicht darüber hinweg täuschen, dass in der Nachkriegszeit bis 1950 die therapeutische Forschung in Deutschland keine wesentliche methodische Weiterentwicklung erfährt, sondern im Gegenteil in einigen Dimensionen eher eine Art Rückschritt.

Bevor mögliche Ursachen für das Stagnieren der methodischen Entwicklung diskutiert werden, bietet es sich an, zu überprüfen, zu welchem Resultat Stoll bei ihrer Beurteilung der Situation der klinisch therapeutischen Forschung für den Zeitraum zwischen 1918 und 1932 kommt und wie sich die Ergebnisse beider Arbeiten im Vergleich zueinander einordnen lassen.

Stoll hat in den Zeitschriften „Deutsche Medizinische Wochenschrift“, „Münchener Medizinische Wochenschrift“, „Klinische Wochenschrift“, „Deutsches Archiv für klinische Medizin“ und „Verhandlungen der Gesellschaft für innere Medizin“ insgesamt 268 therapeutische Studien für den Zeitraum zwischen 1918 und 1932 evaluiert.¹⁶⁶

Das entspricht 18 Studien pro Jahr. Zum Vergleich ergab sich in dem von mir untersuchten Zeitraum ein Durchschnittswert von neun Studien jährlich. Ein Großteil dieser Differenz

¹⁶⁶ Vgl Stoll 2005 S. 99

scheint ein Ausdruck unterschiedlicher methodischer Ansätze in beiden Recherchen darzustellen. Dennoch sollte diskutiert werden, welche möglichen Ursachen dieser erhebliche Unterschied hat.

In beiden Arbeiten ist der Fokus auf die therapeutische Forschung in der inneren Medizin gerichtet. Zwar hat Stoll im Unterschied zu dieser Arbeit auch Studien aus dem Bereich der Pädiatrie mit in die Evaluation einbezogen.¹⁶⁷ Jedoch umfasst die Gesamtheit pädiatrischer Studien gerade einmal 23 Publikationen, was einem Anteil von 9 % an ihrem Gesamtmaterial entspricht. Subtrahiert man die Anzahl pädiatrischer Studien von der Gesamtzahl, so ergibt sich ein Wert von 245 Studien, entsprechend 16 Studien pro Jahr. Auch unter Berücksichtigung dieses Unterschiedes lässt sich die erhebliche Differenz der Anzahl klinischer Studien zwischen beiden Arbeiten nicht erklären.

Die Anzahl und die Titel der in die Recherche einbezogenen Zeitschriften sind in beiden Arbeiten ebenfalls identisch. Die Verteilung der 268 Studien auf die verschiedenen Studientypen ist etwas unterschiedlich. So hat Stoll zum Beispiel 49 % als experimentelle Studien, 25 % als Beobachtungsstudien, 18 % als Fallserien, 0,5 % als Querschnittsstudien und 3,5 % als nicht kategorisierbare Studien identifiziert.¹⁶⁸ Zum Vergleich wurden unter den 158 Studien dieser Arbeit 42 % den experimentellen Studien, 39 % den Beobachtungsstudien, 15 % den Fallserien und 5 % den Querschnittstudien zugeordnet. Auch wenn sich die Ergebnisse bei dem Anteil der Beobachtungsstudien und der Querschnittstudien stärker unterscheiden, scheint doch die in beiden Arbeiten angewandte Methodik zur Identifizierung und Typisierung von therapeutischen Studien ungefähr vergleichbar zu sein. Der Vermutung, dass Stoll bei der Differenzierung zwischen wissenschaftlichen Studien und narrativen Erfahrungsberichten großzügiger zugunsten ersterer verfahren wäre und dadurch mehr Publikationen in die Auswertung miteinbezogen hätte, widerspricht die Beobachtung, dass Stoll etwa 49 % der ausgewerteten Veröffentlichungen als Studie mit einem experimentellen Charakter identifiziert hat und damit 7 % mehr als in der vorliegenden Recherche. Studien mit experimentellem Charakter lassen sich weniger mit unsystematischen narrativen Erfahrungsberichten verwechseln, da für eine experimentelle Versuchsanordnung eine ex-ante Planung, die ein zentrales Differenzierungskriterium zwischen systematischen und unsystematischen Arbeiten darstellt, unverzichtbar ist. Umgekehrt deutet also der Vergleich der Verteilung der Studientypen in beiden Recherchen eher daraufhin, dass in der vorliegenden Arbeit ein etwas großzügigerer Gebrauch der Definition von wissenschaftlichen Studien angewandt

¹⁶⁷ Stoll 2005, S. 28.

¹⁶⁸ Stoll 2005, S. 99.

wurde. Dennoch zeigt sich zwischen beiden Arbeiten ein erheblicher Unterschied in der Anzahl evaluierter Studien zugunsten der von Stoll. Auch die Annahme, dass der zweite Weltkrieg, der in den Zeitraum meiner Beobachtung fällt, für den großen Unterschied verantwortlich sei, ist nicht haltbar, da in der Zeit vor dem Krieg zwischen 1933 und 1938 etwa nur neun Studien jährlich evaluiert worden sind und der Krieg zu Beginn sogar einen eher positiven Einfluss auf die Anzahl therapeutischer Studien hat, wie weiter oben bereits ausgeführt wurde. Diese Beobachtung steht interessanterweise im Gegensatz zur Feststellung von Stoll, die den Effekt des ersten Weltkriegs auf die Häufigkeit therapeutischer Studien als negativ bewertet.¹⁶⁹ Insofern lässt sich diese Differenz mit den zugänglichen Informationen nicht ganz erklären.

Nicht belegbar, aber zumindest diskussionswürdig ist, dass neben methodischen Unterschieden zwischen beiden Literaturrecherchen auch wirtschaftliche oder politische Ereignisse, wie etwa die Weltwirtschaftskrise von 1929 oder die Machtübernahme durch Hitler 1933, tatsächlich Einfluss auf die Anzahl klinischer Studien gehabt haben. Stoll fand heraus, dass zwischen Instituten, die klinische Studien betreuten, und Pharmafirmen oftmals Verbindungen existierten.¹⁷⁰ Auch nach 1932 finden sich Veröffentlichungen, aus denen solche Verbindungen hervorgehen. Oft lässt sich das in den Publikationen verfolgen, indem zum Beispiel einer Firma dafür gedankt wurde, dass dem Krankenhaus oder Institut Medikamente zur Prüfung zur Verfügung gestellt wurden. Man kann vermuten, dass die Durchführung von einigen klinischen Studien wesentlich von der Unterstützung von Firmen abhängig gewesen ist und dass Firmen, die unter wachsenden wirtschaftlichen Problemen gelitten haben, ihre Sponsorentätigkeit im Bereich der klinischen Forschung zunehmend eingeschränkt haben. In dem Material von Stoll lässt sich keine eindeutige Zäsur in der Anzahl klinischer Studien nach 1929 beobachten. Jedoch könnten die Folgen der Krise langfristig gewesen sein. Außerdem könnte ein weiteres Ereignis den Effekt der Weltwirtschaftskrise verstärkt haben.

Die Machtübernahme durch das Nationalsozialistische Regime von 1933 in Deutschland könnte ein weiterer Grund für die deutliche Abnahme klinischer Studien nach 1932 gewesen sein. Folgen der ab 1933 wachsenden Repression und zunehmenden Verfolgung von politischen Gegnern und jüdischen Bürgern in Deutschland waren personelle Veränderungen an wissenschaftlichen und öffentlichen Einrichtungen. Diese Veränderungen haben möglicherweise ebenfalls einen Effekt auf die Durchführung

¹⁶⁹ Vgl. Stoll 2005, S. 94: So beobachtet Stoll in der Zeit zwischen 1914 bis 1918 einen Rückgang der Anzahl klinischer Studien und beschreibt dieses mit dem Begriff „Kriegsloch“.

¹⁷⁰ Stoll 2005, S. 180.

klinischer Studien gehabt, indem zum Beispiel klinische Forscher aus ihren Positionen verdrängt wurden und dadurch ein Verlust wissenschaftlichen Potentials entstanden ist.¹⁷¹ Ein weiterer Erklärungsansatz für die geringe Anzahl ausgewerteter Studien in der Zeit nach 1932 könnte auch die Neuformulierung der Reichsrichtlinien zur Forschung am Menschen von 1931 liefern.¹⁷² Die Neuformulierung der Reichsrichtlinien ist vor allem auch als Reaktion auf den Lübecker Impfskandal von 1930 zu verstehen.¹⁷³ Die Reichsrichtlinien lassen sich im Wesentlichen als eine Modifikation der Preußischen Anweisungen von 1900 auffassen.¹⁷⁴ In einigen Punkten stellten die Richtlinien allerdings eine erhebliche Verschärfung der juristischen Grundlagen für klinische Versuche im Vergleich zu den Preußischen Anweisungen dar. So sei zum Beispiel in den Reichsrichtlinien, die Aufklärung und Einwilligung von Versuchspersonen als Grundvoraussetzung zur Durchführung von klinischen Versuchen neu eingeführt worden. Ferner sei verlangt worden, dass dem Humanexperiment Tierversuche vorausgegangen sein müssten.¹⁷⁵ Auch wenn anzunehmen ist, dass der Effekt der Reichsrichtlinien als insgesamt gering einzuschätzen ist,¹⁷⁶ ist es doch wahrscheinlich, dass ein Teil von Studien, in denen vielleicht aus methodischen oder anderen Gründen ein Vergehen an den Reichsrichtlinien durchgeführt worden ist, nicht publiziert wurde, um einer Strafverfolgung zu entgehen. Die teilweise lückenhafte Darstellung in einigen Publikationen könnte ebenfalls eine Konsequenz aus der Verschärfung der Richtlinien sein. Das könnte sich dadurch erklären, dass Autoren, indem sie rechtswidrige Umstände der Versuche unerwähnt lassen, eine juristische Verfolgung vermeiden wollten. Wahrscheinlich wird die Differenz der Studienanzahl in beiden Recherchen zum Großteil auf methodische Unterschiede bei der Literaturrecherche zurückzuführen sein. Dennoch bleibt zu beachten, dass in diesem Zeitraum auf wirtschaftlicher und politischer Ebene gravierende Veränderungen eingetreten sind, die möglicherweise durch ihre zeitliche Nähe und in Kombination miteinander ebenfalls einen Teil der Differenz mitverantworten. Vergleicht man die Ergebnisse von Stoll bezüglich der methodischen Durchführung mit den Ergebnissen dieser Recherche, muss man bedenken, dass ein solcher Vergleich in einigen Punkten Grenzen hat. Das bezieht sich vor allem auf Merkmale, die, wie oben

¹⁷¹ Vgl. Schmiedeback 1986, S. 196 f.: Schmiedeback beschreibt so zum Beispiel, dass das Gesetz vom 7.4.1933 zur „Wiederherstellung des Berufsbeamtentums“ missbraucht wurde, um jüdisches Personal systematisch zu diskriminieren und aus dem Berufsleben zu verdrängen.

¹⁷² Vgl. Reichsrichtlinien zur Forschung am Menschen 1931 174 f.

¹⁷³ Roelcke 2007, S. 138.

¹⁷⁴ Stoll 2005, S. 156 f.

¹⁷⁵ Stoll 2005, S. 156.

¹⁷⁶ Vgl. Sauerteig 2000, S. 333: Insgesamt ist davon auszugehen, dass die Reichsrichtlinien einen geringen Effekt auf die Forschungspraxis gehabt haben.

erwähnt, einer objektiven Beurteilung schwerer zugänglich sind als andere. Das betrifft insbesondere Fragen nach einer „kritischen“ oder einer „methodischen“ Auseinandersetzung innerhalb der Studien, da deren Beurteilung stark von der individuellen Betrachtung und Bewertung des Untersuchers abhängt. Für andere Merkmale wiederum scheint eine vergleichende Betrachtung einfacher möglich. Stoll konnte in ihrer Recherche in 68 % der 268 Studien einen Kontrollmechanismus auffinden. Eine alternierende Gruppenbildung konnte sie in fünf Publikationen entsprechend einem Anteil von 2 % am Gesamtmaterial feststellen.¹⁷⁷ Zum Vergleich hatten in der vorliegenden Recherche 90 % der Studien einen Kontrollmechanismus und in 14 % wurde ein alternierendes Verfahren bei der Gruppenbildung angewandt. Bezieht man die Anzahl der Studien mit alternierender Gruppenbildung nicht auf die Gesamtheit aller evaluierter Studien, sondern auf die experimenteller Studien, so ergibt sich für Stolls Arbeit ein Wert von 4 % und für diese Arbeit ein Wert von 34 %. Während die relativen Unterschiede in der Anwendung von Kontrollmechanismen im Vergleich zu der Zeit vor 1933 zwar durchaus bemerkenswert sind, lässt sich für den Gebrauch des alternierenden Verfahrens zur Gruppenbildung eine noch umfassendere Veränderung feststellen. Findet sich eine alternierende Gruppenbildung zwischen 1918 und 1932 in nur 4 % aller experimentellen Studien wieder, steigt der Anteil kontinuierlich in der Vorkriegsperiode auf 18 %, in der Kriegsperiode auf 39 % und erreicht in der Nachkriegsperiode schließlich einen Wert von 56 %. Für dieses Merkmal lässt sich mit großer Sicherheit ein Fortschritt dokumentieren, da in diesem Fall eine subjektive Beurteilung dieses Merkmals durch unterschiedliche Untersucher fast ausgeschlossen werden kann. Es zeigt sich, dass sich das alternierende Verfahren zunehmend als Standardmethode bei der Gruppenbildung durchgesetzt hat.

Auch für die Dimension der Ergebnisauswertung deuten sich im Vergleich der Arbeiten Fortschritte an. Als einfache statistische Verfahren definiert Stoll die Bildung von Mittelwerten oder prozentualen Angaben, als komplexe statistische die Anwendung des mittleren Fehlers bei der Ergebnisauswertung.¹⁷⁸ In dieser Arbeit wurde Stolls Definition übernommen und daher ist ein Vergleich dieses Merkmals möglich. Stoll beschreibt, dass in 184 Publikationen eine statistische Auswertung erfolgt sei, entsprechend 69 % am Gesamtmaterial. In 180 Studien (67 %) sei eine einfache statistische Auswertung durchgeführt worden und in vier Studien (2 %) eine komplexe.¹⁷⁹ Im vorliegenden

¹⁷⁷ Stoll 2005, S. 106.

¹⁷⁸ Stoll 2005, S. 110.

¹⁷⁹ Stoll 2005, S. 111.

Material wurde in 112 Studien (71 %) eine statistische Auswertung durchgeführt. In 93 Publikationen (59 %) wurden einfache, in 19 (12 %) wurden komplexere statistische Verfahren eingesetzt. Zwar lässt sich im Vergleich der Zeiträume zwischen 1918 bis 1932 und 1933 bis 1950 kein signifikanter Unterschied in der Gesamtzahl der Studien erkennen, in denen eine statistische Auswertung durchgeführt wurde, wohl aber in dem Verhältnis von Publikationen mit komplexeren statistischen Auswertungen zu der Gesamtzahl an Veröffentlichungen. Noch deutlicher zeigt sich diese Tendenz zur Anwendung von komplexeren Auswertungen, wenn man die Perioden des Zeitraums zwischen 1933 und 1950 in den Vergleich mit einbezieht. Beträgt der Anteil von Studien mit komplexer statistischer Auswertung in der Vorkriegsperiode nur 6 %, steigt dieser Anteil auf 16 % in der Kriegs- beziehungsweise auf 15 % in der Nachkriegsperiode an. Es lässt sich also ein bemerkenswerter Fortschritt bei der Ergebnisauswertung im Vergleich zur der Zeit vor 1933 feststellen, auch wenn dieser in der Nachkriegszeit nicht mehr ganz so deutlich erscheint.

Wie weiter oben bereits festgestellt wurde, zeigt sich in dem Material eine überraschend geringe Anzahl von Studien, in denen eine verblindete Versuchsanordnung oder eine Placebokontrolle umgesetzt wurde. In dem Material konnten gerade einmal vier Studien (ca. 3%), die eine Placebokontrolle angewandt haben, und nur sieben Veröffentlichungen (ca. 4 %), in denen eine Verblindung eingesetzt wurde, gefunden werden. Stoll hat zum Vergleich acht Publikationen (3 %) mit Placebokontrollen identifiziert und 13 (5 %) mit einer unwissentlichen Versuchsanordnung.¹⁸⁰ Es lässt sich für diese beiden Merkmale also keine zunehmende Anwendung in klinischen Studien nach 1933 feststellen. Im Gegenteil verschwinden sie aus der Durchführung klinischer Studien nach 1945 völlig. Interessant wäre es gewesen, zu überprüfen, welcher Disziplin die Studien mit Placebokontrolle oder Verblindung zuzuordnen sind, die Stoll aufgefunden hat. Denn die wenigen von mir gefundenen Studien mit Anwendung dieser methodischen Raritäten befassten sich mit der Erforschung von psychotropen Substanzen und unterstützen die Vermutung, dass die Forschung aus diesem Bereich einen besonderen Stellenwert für die methodische Entwicklung therapeutischer Studien einnimmt. Leider fanden sich in dem Text von Stoll keine Angaben bezüglich der Disziplin dieser Studien, so dass diese Annahme an ihrem Material nicht überprüfbar bleibt.

Ein Problem zeigt sich bei dem Vergleich der Ergebnisse zu den Fragen, ob eine allgemein kritische oder eine methodische Diskussion stattgefunden hat. Die Ergebnisse dazu sind

¹⁸⁰ Stoll 2005, S. 100 ff.

sehr verschieden und deuten daraufhin, dass unterschiedliche Kriterien bei der Bewertung dieser Fragen angewandt wurden. Stoll hat zum Beispiel für 49 % der Studien eine „allgemein kritische“ und für 10 % eine „methodische“ Diskussion beschrieben.¹⁸¹ Zum Vergleich wurde in der vorliegenden Recherche in nur 30 % der Studien eine „allgemein kritische“, dafür aber in 60 % eine „methodische“ Diskussion beobachtet. Da sich beide Merkmale nicht so eindeutig definieren und objektiv abgrenzen lassen wie etwa Fragen zu der Art der angewandten statistischen Auswertung oder der Verblindung, scheinen sich die gravierenden Unterschiede doch am ehesten durch eine unterschiedliche Beurteilung des jeweiligen Untersuchers zu erklären. Insofern scheint ein Vergleich hinsichtlich dieser Kategorien nicht unbedingt sinnvoll zu sein. Ähnliche Ergebnisse liefern allerdings Fragen bezüglich der ethischen Auseinandersetzung innerhalb der Studien, was auf eine Vergleichbarkeit wenigstens hinsichtlich dieses Kriteriums hinweist. Stoll hat in ihrem Material eine ethische Diskussion in 6 % der Publikationen verfolgt,¹⁸² während in der vorliegenden Arbeit dieser Anteil bei etwa 8 % liegt. Man kann hinsichtlich dieses Merkmals keine gravierenden Unterschiede feststellen. Mit einer zu mutigen Interpretation dieser Beobachtung sollte man allerdings sehr zurückhaltend sein, da die historische Dimension der Forschung im Krieg eventuell zu unzulässigen Schlussfolgerungen bezüglich der ethischen Situation führen könnte.

Zusammengefasst stehen vier zentrale Beobachtungen am Ende dieser vergleichenden Betrachtung.

Erstens wurden insgesamt deutlich weniger klinische Studien evaluiert. Wahrscheinlich werden hauptsächlich Unterschiede in der Methodik zwischen beiden Arbeiten dafür verantwortlich sein. Dennoch könnten auch die erwähnten Ereignisse zu dieser Differenz beigetragen haben, wobei sich deren Wirkung nicht exakt quantifizieren lässt.

Zweitens ließ sich zeigen, dass der zweite Weltkrieg im Gegensatz zum ersten offenbar zu einer Zunahme klinischer Studien geführt hat.

Die dritte Erkenntnis ist, dass auf einer qualitativen Ebene eine durchaus positive Entwicklung einsetzt. Auch wenn man berücksichtigt, dass placebokontrollierte und verblindete Studien tendenziell seltener durchgeführt worden sind, so lässt sich doch ein deutlich zunehmender Anteil von Studien mit alternierenden Verfahren und mit statistisch komplexen Methoden am Gesamtmaterial beobachten.

¹⁸¹ Stoll 2005, S. 112.

¹⁸² Stoll 2005, S. 112.

Viertens lässt sich noch feststellen, dass ein Vergleich zwischen beiden Arbeiten auf allen Ebenen nicht möglich ist. Das dürfte in erster Linie an Unterschieden in der Methodik liegen.

Weiter ist zu überlegen, welchen Einfluss Paul Martinis Ideen auf die methodische Gestaltung klinischer Forschung gehabt haben. In etwa 30 % des Gesamtmaterials, werden zentrale Forderungen von Martini umgesetzt. Dabei kann man in etwa 10 % einen direkten Bezug auf Martini feststellen. Der Anteil von Studien, in denen die Vorschläge Martinis implementiert wurden, wächst zwischen der Vor- und der Kriegsperiode deutlich von 18 % auf 35 % an. In der Nachkriegsperiode setzt sich dieses Wachstum nicht fort und der Anteil von Studien mit Berücksichtigung von Martinis Vorschlägen geht leicht zurück, während zumindest der Anteil mit direktem Bezug zu Martini auf 20 % anwächst. Inhaltlich fanden vor allem die Ideen zur Bildung von Kontrollgruppen Anklang. Andere Vorschläge aber, wie etwa die Anwendung statistischer Größen bei der Ergebnisauswertung oder die Versuchsverblindung, wurden deutlich seltener umgesetzt.¹⁸³

In der 1. Auflage der Methodenlehre wird deutlich, woher Martini wichtige Impulse für seine Vorschläge zur Gestaltung klinischer Studien bekommen hat. In der Methodenlehre bezieht er sich mehrfach auf den Schweizer Eugen Bleuler und den Österreicher Julius Wagner-Jauregg, die beide als Psychiater tätig waren. Martini bezeichnet Wagner-Jauregg als Namensgeber für die Simultanmethode.¹⁸⁴ Die Simultanmethode beschreibt ein alternierendes Verfahren zur Gruppenbildung.

Eugen Bleulers Bedenken bei der Beurteilung von therapeutischen Erfolgen greift Martini auf und mahnt zur Zurückhaltung bei der Bewertung der Wirksamkeit von therapeutischen Interventionen.¹⁸⁵ Bleuler selbst hat seine Forderungen an die Umsetzung klinischer Forschung in dem Buch „Das autistisch-undisziplinierte Denken in der Medizin und seine Überwindung“ 1919 zusammengefasst.¹⁸⁶ Dabei stehen im Zentrum seiner Forderungen die Anwendung mathematischer Operationen bei der Ergebnisbeurteilung sowie vergleichende Beobachtungen im Sinne eines alternierenden Versuchsaufbaus.¹⁸⁷ Speziell Bleulers Kritik und seine Vorschläge zur Optimierung klinischer Studien stimmen in den zentralen Forderungen, eine Vergleichsgrundlage zu schaffen und eine statistisch

¹⁸³ Vgl. Winau 1984 S, 100 f.: Diese Ergebnisse stimmen mit der Beobachtung von Winau überein, dass etwa der einfache Blindversuch und die Anwendung von Placebokontrollen erst nach 1950 in nennenswertem Umfang in der klinischen Medizin in Deutschland eingesetzt wurden.

¹⁸⁴ Martini 1932, S. 4.

¹⁸⁵ Martini 1932, S. 38 f.

¹⁸⁶ Bleuler 1919.

¹⁸⁷ Vgl. Bleuler 1919, S. 119 ff.

genügende Auswertung klinischer Studien durchzuführen, mit denen Martinis überein.¹⁸⁸ Insofern kann man Bleuler als einen Vordenker von Martini betrachten. Dass Wagner-Jauregg und Bleuler, die einen erkennbaren Einfluss auf Martini gehabt haben, beide als Psychiater gearbeitet haben, lässt sich durchaus als Hinweis für die wichtige Bedeutung der neurowissenschaftlichen Forschung für die methodische Entwicklung in der klinischen Medizin verstehen. Das deckt sich mit der Beobachtung, dass sämtliche placebokontrollierte Studien und vier von insgesamt sieben verblindeten Studien dem Bereich der Neurowissenschaften zuzuordnen sind. Die Placebokontrolle und die verblindete Versuchsanordnung sind speziell für die Psychiatrie und die Psychologie als sehr wichtige Ergänzungen des methodischen Repertoires zu verstehen. Denn insbesondere in diesem Bereich der klinischen Wissenschaft sind die Ergebnisse stark von der subjektiven Beurteilung des Probanden abhängig und daher für suggestive Einflüsse besonders anfällig. Es lässt sich entsprechend vermuten, dass im Bereich der psychiatrischen und psychologischen Forschung deutlich früher angefangen wurde, den Wert von Placebokontrollen und Verblindungen zu erkennen und in die Versuchsanordnung einzuführen als in der allgemein therapeutischen Forschung. Insofern dürfte die Auseinandersetzung mit psychiatrischen Kollegen für Martini wertvolle Gedankenanstöße gehabt haben, wenn er so zum Beispiel die Verblindung in klinischen Studien fordert. Aus dem Meinungs austausch mit Koller und Heidenhain geht deutlich hervor, dass die Überlegungen in der Methodenlehre zur statistischen Auswertung klinischer Studien umstritten waren. Die Kritik von Koller¹⁸⁹ stimmt aber keineswegs mit der von Heidenhain¹⁹⁰ überein. Umso mehr muss diese Diskussion einen verwirrenden Eindruck auf den Leser hinterlassen haben. Außerdem macht dieser Meinungs austausch deutlich, dass offenbar keine allgemein akzeptierten und etablierten Methoden zur statistischen Auswertung klinischer Studien existierten. Ferner lassen sich in der Stellungnahme Heidenhains Hinweise erkennen, dass auch die Praxis des klinischen Vergleichs wie von Martini gefordert als nicht allgemein anerkannt betrachtet werden darf. Übereinstimmend kommen verschiedene Wissenschaftshistoriker zu dem Urteil, dass Martinis Methodenlehre ein in dem zeitlichen Kontext bemerkenswerten Versuch darstellt, den kontrollierten klinischen Versuch in der medizinischen Forschung zu etablieren und zu standardisieren.¹⁹¹ Dennoch weist die Methodenlehre insbesondere in späteren Auflagen

¹⁸⁸ Vgl. Bleuler 1919.

¹⁸⁹ Vgl. Koller 1934.

¹⁹⁰ Vgl. Heidenhain 1934.

¹⁹¹ Vgl. Kaptchuk 1998, S. 421; vgl. Neuhaus 1984 S. 116.

Schwächen auf. Wesentlicher Schwachpunkt ist dabei, dass Martini in der zweiten Auflage die ausgleichende Alternierung bei der Gruppenbildung beschreibt. Ein Jahr nach der Publikation der zweiten Auflage 1947, wurde bereits die erste randomisierte kontrollierte Studie veröffentlicht. Insofern sind die Vorschläge der zweiten Auflage der Methodenlehre als zu dem Zeitpunkt nicht mehr so fortschrittlich zu verstehen.¹⁹²

Eine weitere Frage dieser Arbeit ist, ob ein Austausch zwischen deutschen und ausländischen Forschern über methodische Fragestellungen stattgefunden hat. Dazu ist zunächst zu bemerken, dass es im Rahmen dieser Recherche schwierig ist, sichere Aussagen über einen möglichen Einfluss ausländischer Autoren auf die methodische Entwicklung in der klinischen Forschung in Deutschland zu machen. Eine Messung des Einflusses auf der Grundlage von zitierten Arbeiten aus dem Ausland in den Studien stellt lediglich eine vage Annäherung an diese Frage dar. Im Material lässt sich feststellen, dass in der Kriegszeit mehr Autoren aus dem Ausland zitiert wurden als in der Vorkriegszeit. Diese Beobachtung sollte allerdings aus zwei Gründen nicht als Beweis missverstanden werden, dass tatsächlich methodische Überlegungen aus dem Ausland ebenfalls in einem größeren Maß wahrgenommen wurden. Erstens war die Publikationspraxis nicht standardisiert, so dass die Literaturverzeichnisse der Studien teilweise lückenhaft und damit nicht zuverlässig waren. Und zweitens finden sich in den wenigsten Veröffentlichungen Hinweise darauf, welche Inhalte der zitierten Literatur zur Kenntnis genommen wurden. Auffällig ist zumindest die Beobachtung, dass sich daraus keine Hinweise auf eine zunehmende Isolation deutscher Forscher während des Krieges finden.

Für die Fragestellung sind auch Veröffentlichungen von ausländischen Autoren in deutschen Zeitschriften, wie die von Poul Bahnsen, aufschlussreich.¹⁹³ An dem Beispiel dieser Veröffentlichung lässt sich verfolgen, wie im Ausland, in diesem Fall in Dänemark, Forschung methodisch umgesetzt wurde. Aus der Studie von Bahnsen geht hervor, dass der Einsatz von Placebokontrollen oder einer verblindeten Versuchsanordnung nicht nur bekannt war, sondern auch in groß angelegten Studien mit weit über 100 Teilnehmern praktiziert wurde, während die in dieser Arbeit evaluierten placebokontrollierten oder verblindeten Studien in einem deutlich kleineren Rahmen stattgefunden haben.¹⁹⁴ Die methodische Diskussion in dieser Studie war besonders detailliert und hatte fast eine Art

¹⁹² Tröhler 1993 S. 73.

¹⁹³ Bahnsen et al 1938.

¹⁹⁴ Vgl. Bahnsen et al 1938 S. 1075.

Fortbildungscharakter für den Leser. Während die methodischen Vorschläge Martinis, die er in der Methodenlehre zusammengefasst hat, im Wesentlichen den klinischen Forschern bekannt gewesen sein dürften, die sich bewusst und gezielt mit diesen Fragestellungen auseinandergesetzt haben, werden Veröffentlichungen in überregionalen Zeitschriften wahrscheinlich eine deutlich größere Anzahl von Klinikern erreicht und dadurch eventuell einen höheren Wirkungsgrad gehabt haben. Dieser Wirkungsgrad einer einzelnen Veröffentlichung und speziell der Mecodrin-Studie dürfte aber wegen der nicht standardisierten und lückenhaften Publikationspraxis kaum zu messen sein.

Die Mecodrin-Studie von Bahnsen ist dabei als das vorläufige Resultat einer methodischen Entwicklung zu verstehen, die in Dänemark Ende des 19. Jahrhundert mit den Arbeiten von Fibiger¹⁹⁵, Heiberg und Finsen¹⁹⁶ einsetzt. Dabei wird speziell der Beitrag Fibigers mit der Studie zur „Serumbehandlung der Diphtherie“ von 1897 als ein wichtiger Meilenstein für die Entwicklung zu randomisierten kontrollierten Studien gewürdigt.¹⁹⁷ Als nicht minder wichtig dürfte auch die Arbeit von Heiberg in dieser Entwicklung einzuordnen sein. Mehr als 30 Jahre bevor Martini seine Forderungen an die klinische Forschung formuliert, finden sich bei Heiberg zentrale Punkte Martinis.¹⁹⁸ Für den skandinavischen Raum und im speziellen für Dänemark lässt sich in methodischer Hinsicht ein hoher Entwicklungsstand dokumentieren, der insbesondere in der Studie von Bahnsen zum Ausdruck kommt.

Im angelsächsischen Raum scheint die methodische Durchführung klinischer Studien in den 30er Jahren des 20. Jahrhundert`s ein mit Deutschland vergleichbares Niveau zu haben. In der Literatur finden sich jedenfalls keine Hinweise auf entscheidende Unterschiede in der methodischen Durchführung klinischer Versuche. Zum Ausdruck kommt diese parallele Entwicklung mit der Definition des Begriffs „klinischer Versuch“ und seinen methodischen Rahmenbedingungen, die unabhängig voneinander durch Paul Martini ab 1932 und Sir Austin Bradford Hill ab 1937 erfolgt ist.¹⁹⁹

Nach Kriegsende werden deutlich weniger klinische Studien veröffentlicht. Das hat im Wesentlichen zwei Ursachen. Zum einen sind medizinische Zeitschriften als Folge des Krieges teilweise über mehr als vier Jahre nicht mehr erschienen. Zum anderen wurden zur Kriegszeit etwa 25 % aller klinischen Studien in militärischen Einrichtungen durchgeführt.

¹⁹⁵ Fibiger 1898.

¹⁹⁶ Finsen 1896.

¹⁹⁷ Lilienfeld 1982, S. 10 f.

¹⁹⁸ Heiberg 1898, S. 16 f.: So fordert Heiberg wie Martini, dass für den statischen Vergleich zwischen verschieden behandelten Gruppen Veränderungen in dem Genus epidemicus unbedingt zu berücksichtigen sind.

¹⁹⁹ Vgl. Neuhaus 1984, S. 116.

In der Vorkriegsperiode war dieser Anteil erheblich kleiner (6 %) und in der Nachkriegszeit finden sich keine Veröffentlichungen zu klinischen Versuchen in solchen Einrichtungen wieder. In der Lazarettmedizin wurden die Versuche methodisch auf der Höhe der Zeit durchgeführt. Inhaltlich wurde in der Kriegszeit sowohl in den Lazaretten wie auch in der zivilen Forschung vor allem an der Erforschung antibiotischer Medikamente gearbeitet, was auf die kriegswichtige Bedeutung von Antibiotika hinweist.²⁰⁰

Abgesehen vielleicht von Feldlazaretten in unmittelbarer Frontnähe fanden sich in anderen militärischen Einrichtungen für diese Fragestellungen aus der Sicht klinischer Forscher sehr gute Voraussetzungen. Die Versuchsteilnehmer waren militärisch diszipliniert und organisiert, wodurch sich Probleme, die zum Beispiel durch eine ungenügende Kontrolle der Lebensumstände der Probanden bestanden, besser beherrschen ließen als in zivilen Krankenhäusern. Außerdem bestand das Kollektiv im Vergleich zu zivilen Einrichtungen aus verhältnismäßig jungen und ansonsten gesunden Probanden, womit zum Beispiel wesentlichen Forderungen von Martini an die Homogenität leichter nachgekommen werden konnte.

Unmittelbar nach dem Krieg scheint die klinische Forschung in Deutschland bis in die 50er Jahre keine wesentlichen Impulse für die methodische Entwicklung mehr empfangen zu haben und scheint auf dem Niveau der Kriegszeit zu stagnieren.²⁰¹ Auch wenn keine sicheren Aussagen über die genauen Gründe dafür gemacht werden können, scheint doch der Krieg eine wichtige Rolle gespielt zu haben. Der im Krieg gewachsene Druck, antibiotisch wirksame Medikamente möglichst zügig zu entwickeln und zur Verfügung zu stellen, könnte die Effizienz in der Praxis der Versuchsdurchführung gefördert haben. Anders ausgedrückt: Die zunehmend schlechten Rahmenbedingungen erforderten einen effizienten Umgang mit vorhandenen Ressourcen, der seinerseits eine methodisch präzise Durchführung klinischer Versuche verlangt hat.²⁰² Dabei könnte auch die militärisch geprägte Umgebung im autoritären System des nationalsozialistischen Deutschlands eine wichtige Rolle gespielt haben, in der Form, dass das Schicksal des Individuums dem Schicksal der Gemeinschaft untergeordnet und ethische Bedenken gegen die Praxis der Versuchsdurchführung beiseite geschoben wurden.

²⁰⁰ Vgl. Kanovitch 1997, S. 94 f.

²⁰¹ Winau 1984, S. 100 f.

²⁰² Vgl. Roelcke 2007, S. 143: Die zunehmende Fokussierung auf kriegswichtige Leistungen bei gleichzeitiger Ressourcenknappheit wird von Roelcke als ein wichtiger logistischer Aspekt klinischer Forschung im 2. Weltkrieg angeführt. Formalrechtliche Aspekte seien demgegenüber stärker in den Hintergrund gedrängt worden.

Interessant ist in diesem Zusammenhang auch, dass Martini wesentliche Forderungen wie etwa die unwissentliche Versuchsanordnung und den Vergleich mit einer Kontrollgruppe vor allem auf akut verlaufende Erkrankungen also vor allem auf Infektionserkrankungen bezieht.²⁰³ Diese Aspekte zusammengenommen lassen erkennen, dass sich in der Kriegszeit Voraussetzungen gefunden haben, die einen aus heutiger Sicht positiven Einfluss auf die methodische Entwicklung der Forschung gehabt haben könnten. Dass sich speziell im Nachkriegsdeutschland diese Voraussetzungen geändert zu haben scheinen, könnte auch als Resultat des Aufarbeitungsprozesses von Verbrechen durch nationalsozialistische Ärzte betrachtet werden. Diese These wird insbesondere von Erwin Deutsch vertreten.²⁰⁴ Im Zentrum seiner Argumentation steht seine Schlussfolgerung, dass im Rahmen der Nürnberger Prozesse, in denen auch Ärzte des nationalsozialistischen Regimes angeklagt wurden, die klinische Forschung einen schweren Ansehensverlust erlitten habe. Dieser Imageschaden sei im Wesentlichen eine Folge des Prozesses gewesen, da in ihm nicht ausreichend zwischen wissenschaftlicher Forschung auf der einen und grausamen Versuchen am Menschen ohne wissenschaftlichen Anspruch auf der anderen Seite differenziert worden sei. Ein Effekt dieser Vermengung könnte schließlich dazu beigetragen haben, dass Versuche am Menschen generell stärker abgelehnt wurden. Im Besonderen könnte das für Experimente gegolten haben, in denen die Teilnehmer bei Zuweisung zu verschiedenen Behandlungsarmen scheinbar der Willkür des Forschers oder des Zufalls ausgesetzt worden seien. Der öffentliche Aufarbeitungsprozess in den „Nürnberger Ärzteprozessen“ wird zwar als insgesamt schleppend beschrieben.²⁰⁵ Jedoch ist nicht auszuschließen, dass die Praxis der Versuchsdurchführung von den Resultaten dieses Prozesses früher erfasst wurde.

Mit dem Imageschaden der medizinischen Forschung einhergehend scheint auch der Vertrauensverlust unter Teilen der Bevölkerung in die deutsche Ärzteschaft zu sein. Von verschiedenen Historikern wird die Ansicht vertreten, dass der ärztliche Berufsstand im nationalsozialistischen Deutschland in seiner Gesamtheit als privilegiert gegolten hat und darüber hinaus in einem überdurchschnittlichen Maß politisch organisiert war.²⁰⁶ Ihr kam neben der Arbeit als Mediziner auch ein ideologisch geprägter Auftrag zu, nämlich den „Volkskörper“ zu heilen.²⁰⁷

²⁰³ Martini 1932, S. 4.

²⁰⁴ Vgl. Deutsch 1997, S. 107 f.

²⁰⁵ Schmidt/Frewer 2007, S. 38.

²⁰⁶ Schmiedebach 1987, S. 194 f.

²⁰⁷ Schmiedebach 1987, S. 196.

Vor dem Hintergrund des Generalverdachts auf Komplizenschaft der Ärzte mit dem Regime ist auch der Aufruf zu verstehen, mit dem Thomas Gerst seinen Aufsatz über den Nürnberger Ärzteprozess einleitet und der 1945 auf der Titelseite der Frankfurter Rundschau abgedruckt war: „Ärzte! Ihr müsst das Vertrauen des Volkes wiedergewinnen!“²⁰⁸ Der Imageschaden auf der einen und das mangelnde Vertrauen zu den Ärzten auf der anderen Seite könnte dem Prozess der methodischen Weiterentwicklung in Nachkriegsdeutschland im Weg gestanden hat. Als einen Versuch, das Vertrauen in den ärztlichen Berufsstand und die klinische Forschung wiederherzustellen, ist auch die Kodifizierung ethischer Normen für wissenschaftliche Studien von Nürnberg zu verstehen.²⁰⁹ Zwar wurde in Deutschland als einem der ersten Länder weltweit 1931 der juristische Rahmen für klinische Versuche abgesteckt.²¹⁰ Dieser hat allerdings nicht die Gräueltaten im Namen der medizinischen Forschung in Deutschland verhindern können. Der Nürnberger Kodex von 1946 stellte eine juristisch wirksame Einschränkung der Autonomie der klinischen Forschung dar. Ziel dieser Normierung war zum einen, Gräueltaten im Namen medizinischer Forschung künftig zu verhindern, und zum anderen, Vertrauen in die Medizin mit ihren verschiedenen Akteuren wiederherzustellen.

Greift man die eingangs formulierten Fragen auf, so kann man in dem Zeitraum zwischen 1933 und 1950 keine anhaltende und umfassende Weiterentwicklung im methodischen Sinn in der klinischen Forschung erkennen. Deuten sich im Vergleich zu der Zeit vor 1933 deutliche Fortschritte an, zeigt sich eine Zäsur in dieser Entwicklung vor allem nach Kriegsende.

Die Enttäuschung Martinis über den methodischen Zustand der klinischen Forschung lässt sich allerdings nur zum Teil verstehen. Das Vorwort zu der 2. Auflage ist 1945 geschrieben worden, also 2 Jahre vor der eigentlichen Veröffentlichung. Ein Fortschritt in der methodischen Durchführung klinischer Forschung lässt sich durchaus in der Zeit zwischen 1933 und 1945 beobachten. Insofern scheint die Beurteilung Martinis etwas unverständlich zu sein. Für die Zeit nach 1945 bis 1950 kann allerdings von keinem spürbaren Fortschritt mehr ausgegangen werden und man kann Martinis Kritik als berechtigt nachvollziehen.

²⁰⁸ Gerst 1994, S. 1

²⁰⁹ Schmidt/Frewer 2007, S. 64

²¹⁰ Lepicard 1997, S. 66

7. Zusammenfassung

Die Methodik klinischer Forschung im frühen 20. Jahrhundert

In der vorliegenden Arbeit wurden methodische Aspekte klinischer Forschung in der Zeit zwischen 1933 und 1950 in Deutschland untersucht. Weiter wurde überprüft, ob zu der Zeit vor 1933 methodische Entwicklungen beobachtbar sind. Außerdem wurde auf mögliche Einflüsse auf die methodische Umsetzung klinischer Forschung eingegangen. Vergleicht man die Zeiträume 1918 – 1932 und 1933 - 1950 lässt sich in qualitativer Hinsicht eine durchaus positive Entwicklung beobachten. Das lässt sich so zum Beispiel mit großer Sicherheit für die Anwendung von Vorläuferverfahren randomisierter kontrollierter Studien behaupten.

Betrachtet man den Zeitraum zwischen 1933 und 1950 genauer, so lässt sich keine dauerhafte methodische Weiterentwicklung in der klinischen Forschung feststellen. Bei differenzierter Betrachtung sind punktuelle Fortschritte insbesondere zwischen den Perioden 1933 – 1938 und 1939 – 1945 zu beobachten. Allerdings setzt sich diese Entwicklung in der Nachkriegszeit nicht fort.

Einen fortschrittlichen Beitrag zur Normierung der Methodik klinischer Forschung hat Paul Martini geleistet. Martinis Methodenlehre zeichnet sich durch sein für damalige Verhältnisse einmaliges didaktisches Konzept aus und füllte eine bis dahin offene Lücke. Dennoch lässt sich ein umfassender Einfluss der Methodenlehre auf die klinische Forschung vor allem nach 1945 nicht beobachten.

Mögliche Gründe für das Stagnieren der methodischen Entwicklung sind unter anderem auch in politischen und gesellschaftlichen Veränderungen nach 1945 zu suchen. So ließ sich zum Beispiel nachweisen, dass während des Krieges ein großer Teil klinischer Forschung im militärischen Umfeld durchgeführt wurde. Durch Auflösung der militärischen Strukturen der Wehrmacht, fielen somit wichtige Einrichtungen für die Forschung aus.

Inwiefern diese Veränderungen und die durch die Nürnberger Ärzteprozesse hervorgerufene Debatte um ethische Grundlagen von Menschenversuchen die Methodik der klinischen Forschung nach 1945 beeinflusst haben, lässt sich nicht sicher abschätzen. Die Ergebnisse der vorliegenden Arbeit weisen zumindest auf einen entsprechenden Zusammenhang hin. Vor diesem Hintergrund könnte man Martinis Methodenlehre als Opfer der politischen Veränderungen seiner Zeit verstehen.

8. Anhang

Literaturverzeichnis

Ach, N, Fürstenberg, H. E.: Über die Beeinflussung der geistigen Leistungsfähigkeit durch Recresal., Münchner Medizinische Wochenschrift (1937) S. 290 – 292

Arzneimittelgesetz in der Fassung der Bekanntmachung vom 12. Dezember 2005 (BGBl. I S. 3394), das zuletzt durch Artikel 1 des Gesetzes vom 17. Juli 2009 (BGBl. I S. 1990) geändert worden ist.

Bahnsen, Poul, Jacobsen, Erich, Thesleff, Harriet: Studien über Wirkung von β -Phenyl-Isopropylaminsulfat (Mecodrin) auf normale Menschen, Klinische Wochenschrift (1938) S. 1074 - 1078

Bingel, Adolf: Wirkt das Diphtherieheilserum bei der menschlichen Diphtheriekrankheit spezifisch durch seinen Antitoxingehalt oder unspezifisch?, Deutsche Medizinische Wochenschrift (1949), S. 101 – 103

Bleuler, Eugen: Das autistisch undisziplinierte Denken in der Medizin und seine Überwindung, Julius Springer Verlag, Berlin 1919, 1922.

Breidert, Matthias, Hofbauer, Karl: Placebo: Missverständnisse und Vorurteile, Deutsches Ärzteblatt (2009), S. 751 - 755

Dennig, H., Hoeffler, Hella: Zur Behandlung der Poliomyelitis mit Rekonvaleszentenserum, Münchner Medizinische Wochenschrift (1933), S. 1367 - 1368

Deutsch, Erwin: Der Nürnberger Kodex Das Strafverfahren gegen Mediziner, die zehn Prinzipien und die bleibende Bedeutung des Nürnberger Kodex, in: Ulrich Tröhler / Stella Reiter-Theil (Hrsg.): Ethik und Medizin, Wallstein 1997, S. 103 – 114.

Dietrich, S, Schwiegk, H.: Ein neues Digitalis lanata Glykosid in der Kreislauftherapie, Münchner Medizinische Wochenschrift (1933) S. 563 - 564

Eigler, Gunter: Politeia, in: Gunter Eigler (Hrsg.): Platon Werke Band 4 (1970),
Wissenschaftliche Buchgesellschaft (1970) S. 556 - 557

Elkeles, Barbara: Die Therapie in den Pariser Spitälern und die numerische Methode, in
Barbara Elkeles (Hrsg.): Der moralische Diskurs über das medizinische
Menschenexperiment im 19. Jahrhundert, Gustav Fischer 1999 S.

Elkeles, Barbara: The german debate on human experimentation between 1880 and 1914,
in Volker Roelcke/Giovanni Maio (Hrsg.): Twentieth century ethics of human subjects
research, Franz Steiner Verlag 2004, S. 19 – 33.

Fibiger, Johannes: Om Serumbehandling af Difteri, Hospitalstidende 1898, 309 – 325
(Englische Übersetzung, James Lind Library)

Finsen, Niels: Om Anvendelse I Medicinen af koncentrerede kemiske Lysstraaler,
Gyldendalske Boghandels Forlag, Kjøbenhavn 1896 (Englische Übersetzung James Lind
Library)

Fletcher, Robert, Fletcher Suzanne, Wagner, Edward: Klinische Epidemiologie.
Grundlagen und Anwendung, Dt. Ausgabe, adaptiert und herausgegeben von Johannes
Haerting und Christoph Rink (Übersetzung: Erika Eiselt), Ullstein Medical, Wiesbaden
1999.

Freyseng, Becker: Vergleichend therapeutische Untersuchung über die Wirkung von
Nerium Oleander auf den gestörten Kreislauf, Klinische Wochenschrift 17 (1938) S. 815 –
820.

Gerst, Thomas: Der Auftrag der Ärztekammern an Alexander Mitscherlich zur
Beobachtung und Dokumentation des Prozessverlaufs, Deutsches Ärzteblatt 91, Heft 22/23
(1994).

Heiberg, Poul: Studier over den statistiske Undersøgelsesmetode som Hjaelpemiddel ved
terapeutiske Undersøgelser, Bibliotek for laeger 89, (1897); Englische Übersetzung

Heidenhain, L: Die Erfolgsbeurteilung therapeutischer Untersuchungen (Richtigstellung), Klinische Wochenschrift (1934), S. 1360 – 1361.

Hess, Otto: Zur Therapie schwerer Diphtherie, Deutsche Medizinische Wochenschrift, 90 (1943) S. 716 – 718.

Kanovitch, Bernard: Die medizinische Experimente in den Konzentrationslagern des Nationalsozialismus, in: Ulrich Tröhler / Stella Reiter-Theil (Hrsg.): Ethik und Medizin 1947 – 1997, Wallstein 1997, S. 91 – 101.

Kaptchuk, Ted J.: Intentional Ignorance: A History of Blind Assessment and Placebo Controls in Medicine, Bulletin of the History of Medicine 72 (1998) 389 – 433.

Kisskalt, Karl: Theorie und Praxis der medizinischen Forschung, J. F. Lehmanns Verlag, München – Berlin 1942

Koller, Siegfried: Die Erfolgsbeurteilung therapeutischer Untersuchungen, Klinische Wochenschrift (1934) S. 1154 – 1155.

Lepicard, Etienne: Ethisches Verhalten und ethische Normen vor 1947, in: Ulrich Tröhler/Stella Reiter-Theil (Hrsg.): Ethik und Medizin 1947 – 1997, Wallstein 1997 S. 61 – 74.

Lind, James: A treatise of the scuvy, Kincaid & Donsaldon, Edinburgh 1753.

Lilienfeld, Abraham M: Ceteris paribus: The evolution of the clinical trial, Bulletin of the history of medicine 56 (1982) 1 – 18.

Loudon Irvine: The use of historical controls and concurrent controls to assess the effects of sulphonamides, 1936 – 1945, The James Lind Library (2002)

Löhner, Georg: Die homöopathischen Kochsalzversuche zu Nürnberg in: Georg Löhner (Hrsg.) Von einer Gesellschaft wahrheitsliebender Männer (1835) Nürnberg

Louis, PCA: Recherches sur les effets de la saignée dans quelques maladies inflammatoires et sur l'action de l'émétique et des vésicatoires dans la pneumonie. (Untersuchung über den Nutzen der Aderlasstherapie bei inflammatorischen Erkrankungen und der Pneumonie) Paris: Librairie de l'Académie royale de médecine (1835).

Martini, Paul: Die Erfolgsbeurteilung therapeutischer Untersuchung, Klinische Wochenschrift (1934), S. 872 – 875 , 907 - 912

Martini, Paul: Methodenlehre der therapeutischen Untersuchung, Springer, Berlin 1932. Ab der 2. Auflage unter dem Titel: Methodenlehre der Therapeutisch-Klinischen Forschung, 1947, 1953, 1968 (ab der 4. Auflage mit G. Oberhoffer und E. Welte).

Neuhaus, Günter: Versuche mit kranken Menschen – der kontrollierte klinische Versuch, in: Hanfried Helmchen/Rolf Winau (Hrsg.): Versuche mit Menschen, de Gruyter 1984 S. 108 – 132.

Nissen, N: Erfahrungen mit der Serumtherapie der kruppösen Pneumonie, Deutsche Medizinische Wochenschrift, (1938) S. 221 – 225

Raettig: Kombinierte Fleckfieberbehandlung mit Rekonvaleszentenserum und Eigenblut, Klinische Wochenschrift, 22 (1943), S. 560 - 563

Raspe, Heiner: Zur aktuellen deutschen Diskussion um die Evidenz-basierte Medizin: Brennpunkte, Skotome, divergierende Wertsetzungen, in: Zeitschrift der ärztlichen Fortbildung und Qualitätssicherung, Urban Fischer 97 (2003), S. 689 – 694.

Reichsrichtlinien zur Forschung am Menschen, Rundschreiben des Reichsministers des Inneren, 28 Februar 1931, nach: Reichsgesundheitsblatt 6 (1931) 174 f.

Roelcke, Volker: Forschung am Menschen im 20. Jahrhundert: Historische Kenntnisse – ethische Implikationen, in: Ulf Schmidt, Andreas Frewer (Hrsg.): Standards der Forschung (2007) S. 135 – 150

Rothman, David J.: Der Nürnberger Kodex im Licht früherer Prinzipien und Praktiken im Bereich der Humanexperimente, in: Ulrich Tröhler / Stella Reiter-Theil: Ethik und Medizin 1947 – 1997, Wallstein 1997 S. 75 – 88.

Sauerteig, Lutz: Ethische Richtlinien, Patientenrechte und ärztliches Verhalten bei der Arzneimittelerprobung (1892 – 1931), Medizinhistorisches Journal 35 (2000) S. 303 - 334

Schmidt, Ulf, Frewer, Andreas: Nuremberg Code of Medical Ethics – Geschichte und Ethik des Ärzteprozesses, in Andreas Frewer, Ulf Schmidt (Hrsg.): Standards der Forschung, Peter Lang (2007) S. 37 – 71.

Schmiedebach, Heinz-Peter: Der Arzt als Gesundheitsoffizier – die systematische Militarisierung der Medizin von 1933 bis zum Zweiten Weltkrieg, in: Johanna Bleker /Heinz-Peter Schmiedebach: Medizin und Krieg, Fischer 1987, S. 191 - 208

Seydel, Hiltrud: Statistik in der Medizin. Ein Entwurf zu ihrer Geschichte, Wachholtz, Neumünster 1976. (Kieler Beiträge zur Geschichte der Medizin und Pharmazie, Bd 15).

Sittsen, A.: Auswertung und Bewertung klinischer Ergebnisse, Münchner Medizinische Wochenschrift (1933) Bd. ,S. 1738 - 1739

Stangemann: Zur Frage der Pneumoniebehandlung mit Eubasinum und dessen Nebenwirkungen, Deutsche Medizinische Wochenschrift (1941) Bd 67, S. 983 - 986

Stoll, Susane: Evidence based Medicine – Gibt es eine Deutsche Vorgeschichte?, Med. Diss., Lübeck, 2005.

Tröhler, Ullrich: Die therapeutische „Erfahrung“ – Geschichte ihrer Bewertung zwischen subjektiv sicherem Wissen und objektiv wahrscheinlichen Kenntnissen, in: Johannes Köbberling (Hrsg.): Die Wissenschaft in der Medizin 1993 S. 65 – 81.

Winau, Rolf: Vom kasuistischen Behandlungsversuch zum kontrollierten klinischen Versuch, in: Hanfried Helmchen/Rolf Winau (Hrsg.): Versuche mit Menschen, de Gruyter 1984, S. 83 – 107.

Abbildungsverzeichnis

Abbildung I:	Verteilung der Studien auf Zeitschriften	S. 50
Abbildung II:	Zeitliche Verteilung der Studien	S. 51
Abbildung III:	Häufigkeit von Studientypen	S. 53
Abbildung IV:	Verwendete Kontrollmechanismen	S. 54
Abbildung V:	Verwendung von Ein- und Ausschlusskriterien	S. 55
Abbildung VI:	Gruppenbildung	S. 57
Abbildung VII:	Studien mit Angaben zu Störgrößen	S. 59
Abbildung VIII:	Qualitative Ergebnisdarstellung	S. 61
Abbildung IX:	Inhaltliche Auseinandersetzung	S. 63
Abbildung X:	Zusammenfassung	S. 64
Abbildung XI:	Verteilung der Studien auf Perioden	S. 66
Abbildung XII:	Verteilung Studientypen auf Perioden	S. 68
Abbildung XIII:	Kontrollmechanismen verteilt auf Perioden	S. 69
Abbildung XIV:	Diskussion der Methodik in den Perioden	S. 72
Abbildung XV:	Vergleich zu Stoll	S. 76
Abbildung XVI:	Wahrnehmung von Martini und seiner Methodenlehre	S. 88
Abbildung XVII:	Verweise auf Arbeiten aus dem Ausland	S. 99
Abbildung XVIII:	Überblick Auswertung militärischer Studien	S. 103
Abbildung XIX:	Inhaltliche Differenzierung der Studien	S. 104

Abkürzungsverzeichnis

DMW = Deutsche Medizinische Wochenschrift
MMW = Münchner Medizinische Wochenschrift
kliWo = Klinische Wochenschrift
dtGes = Verhandlungen der Deutschen Gesellschaft für innere Medizin
dtArchiv = Deutsches Archiv für klinische Medizin
MedZeit = Medizinische Zeitschrift
histor.K. = Historische Kontrolle
Vorbeob. = Vorbeobachtung
Nachbeob. = Nachbeobachtung
Vor+Nachbeob. = Vor- und Nachbeobachtung
Auslass. = Auslassversuch
Verumvgl. = Verumvergleich
Placebovgl. = Placebovergleich
Verum+Placebovgl. = Verum- und Placebovergleich
K-Mechanismus = Kontrollmechanismus
K-Gruppen = Kontrollgruppen
zusätzliche Med. = Zusätzliche Medikamentengabe
RCT = randomized controlled trial = randomisierte kontrollierte Studie

Deutsches Reich
Rundschreiben des Reichsministers des Innern,
betr. Richtlinien für neuartige Heilbehandlung und für die
Vornahme wissenschaftlicher Versuche am Menschen.

Vom 28. Februar 1931.

Der Reichsgesundheitsrat hat besonderen Wert darauf gelegt, Vorsorge zu treffen, daß alle Ärzte von den nachstehenden Richtlinien Kenntnis erhalten und nahm von diesem Gesichtspunkte aus einstimmig eine EntschlieÙung an, wonach alle in Anstalten der geschlossenen und offenen Krankenbehandlung oder Krankenfürsorge tätigen Ärzte auf die Beachtung dieser Richtlinien bei ihrem Eintritt unterschriftlich verpflichtet werden sollten.

1. Die ärztliche Wissenschaft kann, wenn sie nicht zum Stillstand -kommen soll, nicht darauf verzichten, in geeigneten Fällen eine Heilbehandlung mit neuen, noch nicht ausreichend erprobten Mitteln und Verfahren einzuleiten. Ebensowenig kann sie wissenschaftliche Versuche am Menschen als solche völlig entbehren, da sonst die Fortschritte in der Erkennung, der Heilung und der Verhütung von Erkrankungen gehemmt oder sogar ausgeschlossen würden.
Den hiernach den Ärzten einzuräumenden Rechten steht die besondere Pflicht des Arztes gegenüber, sich der großen Verantwortung für Leben und Gesundheit jedes einzelnen, den er neuartig behandelt oder an dem er einen Versuch vornimmt, stets bewußt zu bleiben.
2. Unter neuartiger Heilbehandlung im Sinne dieser Richtlinien sind Eingriffe und Behandlungsweisen am Menschen zu verstehen, die der Heilbehandlung dienen, also in einem bestimmten einzelnen Behandlungsfall zur Erkennung, Heilung oder Verhütung einer Krankheit oder eines Leidens oder zur Beseitigung eines körperlichen Mangels vorgenommen werden, obwohl ihre Auswirkungen und Folgen auf Grund der bisherigen Erfahrungen noch nicht ausreichend zu übersehen sind.
3. Unter wissenschaftlichen Versuchen im Sinne dieser Richtlinien sind Eingriffe und Behandlungsweisen am Menschen zu verstehen, die zu Forschungszwecken vorgenommen werden, ohne der Heilbehandlung im einzelnen Falle zu dienen, und

deren Auswirkungen und Folgen auf Grund der bisherigen Erfahrungen noch nicht ausreichend zu übersehen sind.

4. Jede neuartige Behandlung muß in ihrer Begründung und ihrer Durchführung mit den Grundsätzen der ärztlichen Ethik und den Regeln der ärztlichen Kunst und Wissenschaft in Einklang stehen.

Stets ist sorgfältig zu prüfen und abzuwägen, ob die Schäden, die etwa entstehen können, zu dem erwarteten Nutzen im richtigen Verhältnis stehen.

Eine neuartige Heilbehandlung darf nur vorgenommen werden, wenn sie vorher, soweit als möglich, im Tierversuch geprüft worden ist.

5. Eine neuartige Heilbehandlung darf nur vorgenommen werden, nachdem die betreffende Person oder ihr gesetzlicher Vertreter auf Grund einer vorangegangenen zweckentsprechenden Belehrung sich in unzweideutiger Weise mit der Vornahme einverstanden erklärt hat.

Fehlt die Einwilligung, so darf eine neuartige Heilbehandlung nur dann eingeleitet werden, wenn es sich um eine unaufschiebbare Maßnahme zur Erhaltung des Lebens oder zur Verhütung schwerer Gesundheitsschäden handelt und eine vorherige Einholung der Einwilligung nach der Lage der Verhältnisse nicht möglich war.

6. Die Frage der Anwendung einer neuartigen Heilbehandlung ist mit ganz besonderer Sorgfalt zu prüfen, wenn es sich um Kinder und jugendliche Personen unter 18 Jahren handelt.

7. Die ärztliche Ethik verwirft jede Ausnutzung einer sozialen Notlage für die Vornahme einer neuartigen Heilbehandlung.

8. Bei neuartiger Heilbehandlung mit lebenden Mikroorganismen, insbesondere mit lebenden Krankheitserregern, ist erhöhte Vorsicht geboten. Sie ist nur dann als zulässig zu erachten, wenn eine relative Unschädlichkeit des Verfahrens anzunehmen und auf andere Weise die Erzielung eines entsprechenden Nutzens unter den gegebenen Verhältnissen nicht zu erwarten ist.

9. In Kliniken, Polikliniken, in Krankenhäusern oder sonstigen Anstalten zur Krankenbehandlung und Krankenfürsorge darf eine neuartige Heilbehandlung nur vom leitenden Arzt selbst oder in seinem ausdrücklichen Auftrag und unter seiner vollen Verantwortung von einem anderen Arzt ausgeführt werden.

10. Über jede neuartige Heilbehandlung ist eine Aufzeichnung zu fertigen, aus der der Zweck der Maßnahme, ihre Begründung und die Art ihrer Durchführung ersichtlich

- sind. Insbesondere muß ein Vermerk darüber vorhanden sein, daß die betreffende Person oder erforderlichenfalls ihr gesetzlicher Vertreter vorher zweckentsprechend belehrt worden ist und die Zustimmung gegeben hat.
11. Die Veröffentlichung der Ergebnisse einer neuartigen Heilbehandlung muß in einer Form erfolgen, die der gebotenen Achtung vor dem Kranken und den Geboten der Menschlichkeit in jeder Form Rechnung trägt.
 12. Die Nummern 4 bis 11 dieser Richtlinien gelten entsprechend für wissenschaftliche Versuche (Nr. 3). Außerdem gilt für solche Versuche folgendes:
 - a. Die Vornahme eines Versuchs ist bei fehlender Einwilligung unter allen Umständen unzulässig.
 - b. Jeder Versuch an Menschen ist zu verwerfen, der durch den Versuch am Tier ersetzt werden kann. Ein Versuch am Menschen darf erst vorgenommen werden, wenn zuvor alle Unterlagen beschafft worden sind, die zu einer Klärung und Sicherung mit den der Wissenschaft zur Verfügung stehenden biologischen Methoden des Laboratoriumsversuchs und des Tierexperiments gewonnen werden können. Unter diesen Voraussetzungen verbietet sich jedes grund- und planlose Experimentieren am Menschen von selbst.
 - c. *Versuche an* Kindern oder jugendlichen Personen unter 18 Jahren sind unstatthaft, wenn sie die Versuchsperson auch nur im geringsten gefährden.
 - d. Versuche an Sterbenden sind mit den Grundsätzen der ärztlichen Ethik unvereinbar und daher unzulässig.
 - 13 Wenn man somit von der Ärzteschaft und insbesondere von den verantwortlichen Leitern der Krankenanstalten erwarten darf, daß sie sich von einem starken Verantwortungsgefühl gegenüber den ihnen anvertrauten Kranken leiten lassen, so wird man doch auch bei ihnen diejenige Verantwortungsfreudigkeit nicht entbehren wollen, die auf neuen Wegen den Kranken Erleichterung, Besserung, Schutz oder Heilung zu schaffen sucht, wenn die bisher bekannten Mittel nach ihrer ärztlichen Überzeugung zu versagen drohen.
 - 14 Schon im akademischen Unterricht soll bei jeder geeigneten Gelegenheit auf die besonderen Pflichten hingewiesen werden, die dem Arzt bei der Vornahme einer neuen Heilbehandlung oder eines wissenschaftlichen Versuchs sowie auch bei der Veröffentlichung ihrer Ergebnisse obliegen."

Nürnberger Kodex Fassung vom 20.08.1947

1. Die freiwillige Zustimmung der Versuchsperson ist unbedingt erforderlich. Das heißt, dass der Betreffende die gesetzmäßige Fähigkeit haben muss, seine Einwilligung zu geben; in der Lage sein muss, eine freie Entscheidung zu treffen, unbeeinflusst durch Gewalt, Betrug, List, Druck, Vortäuschung oder irgendeine andere Form der Beeinflussung oder des Zwanges; und genügend Kenntnis von und Einsicht in die Bestandteile des betreffenden Gebietes haben muss, um eine verständnisvolle und aufgeklärte Entscheidung treffen zu können. Diese letzte Bedingung machte es notwendig, dass der Versuchsperson vor der Annahme ihrer bejahenden Entscheidung das Wesen, die Länge und der Zweck des Versuches klargemacht werden; sowie die Methode und die Mittel, welche angewendet werden sollen, alle Unannehmlichkeiten und Gefahren, welche mit Fug zu erwarten sind, und die Folgen für ihre Gesundheit oder ihre Person, welche sich aus der Teilnahme ergeben mögen.

Die Pflicht und Verantwortlichkeit, den Wert der Zustimmung festzustellen, obliegt jedem, der den Versuch anordnet, leitet oder ihn durchführt. Dies sind persönliche Pflichten und persönliche Verantwortungen, welche nicht ungestraft auf andere übertragen werden können.

2. Der Versuch muss derart sein, dass fruchtbare Ergebnisse für das Wohl der Gesellschaft zu erwarten sind, welche nicht durch andere Forschungsmittel oder Methoden zu erlangen sind und welche ihrem Wesen nach nicht willkürlich und unnötig sind.

3. Der Versuch ist so zu planen und auf Ergebnissen von Tierversuchen und einer Kenntnis des Wesens der Krankheit oder des sonstigen Problems aufzubauen, dass die vermutlichen Ergebnisse die Ausführung des Versuchs rechtfertigen werden.

4. Der Versuch ist so durchzuführen, dass alle unnötigen körperlichen und geistigen Leiden und Verletzungen vermieden werden.

5. Kein Versuch darf durchgeführt werden, wenn a priori ein Grund besteht für die Annahme, dass der Tod oder ein dauernder, körperlicher Schaden eintreten wird, mit der Ausnahme vielleicht jener Versuche, bei welchen Versuchsleiter gleichzeitig als Versuchsperson dienen.

6. Das Gefahrenmoment darf niemals die Grenzen überschreiten, welche sich aus der humanitären Bedeutung des zu lösenden Problems ergeben.
7. Angemessene Vorbereitungen sind zu machen und ausreichende Vorkehrungen zu treffen, um die Versuchspersonen gegen selbst die geringste Möglichkeit der Verletzung, der bleibenden gesundheitlichen Schädigungen oder des Todes zu schützen.
8. Der Versuch darf nur von wissenschaftlich geschulten Personen durchgeführt werden. Die größte Geschicklichkeit und die größte Vorsicht müssen in allen Stufen des Versuches von denjenigen angewandt werden, die den Versuch leiten oder durchführen.
9. Während des Versuches muss der Versuchsperson freigestellt bleiben, den Versuch zu beenden, wenn sie körperlich oder geistig den Punkt erreicht hat, an dem ihr seine Fortsetzung unmöglich erscheint.
10. Im Verlauf des Versuchs muss der Versuchsleiter jederzeit bereit bleiben, den Versuch einzustellen, wenn er bei Anwendung des von ihm geforderten guten Glaubens, besonderer Geschicklichkeit und Sorgfalt des Urteils Grund hat anzunehmen, dass eine Fortsetzung des Versuches eine Verletzung, eine bleibende gesundheitliche Schädigung oder den Tod der Versuchsperson herbeiführen könnte.

Gesetz über den Verkehr mit Arzneimitteln

Ausfertigungsdatum: 24.08.1976

§ 4 Sonstige Begriffsbestimmungen

(23) Klinische Prüfung bei Menschen ist jede am Menschen durchgeführte Untersuchung, die dazu bestimmt ist, klinische oder pharmakologische Wirkungen von Arzneimitteln zu erforschen oder nachzuweisen oder Nebenwirkungen festzustellen oder die Resorption, die Verteilung, den Stoffwechsel oder die Ausscheidung zu untersuchen, mit dem Ziel, sich von der Unbedenklichkeit oder Wirksamkeit der Arzneimittel zu überzeugen. Satz 1 gilt nicht für eine Untersuchung, die eine nichtinterventionelle Prüfung ist.

Nichtinterventionelle Prüfung ist eine Untersuchung, in deren Rahmen Erkenntnisse aus der Behandlung von Personen mit Arzneimitteln anhand epidemiologischer Methoden analysiert werden; dabei folgt die Behandlung einschließlich der Diagnose und Überwachung nicht einem vorab festgelegten Prüfplan, sondern ausschließlich der ärztlichen Praxis; soweit es sich um ein zulassungspflichtiges oder nach § 21a Absatz 1 genehmigungspflichtiges Arzneimittel handelt, erfolgt dies ferner gemäß den in der Zulassung oder der Genehmigung festgelegten Angaben für seine Anwendung.

(25) Prüfer ist in der Regel ein für die Durchführung der klinischen Prüfung bei Menschen in einer Prüfstelle verantwortlicher Arzt oder in begründeten Ausnahmefällen eine andere Person, deren Beruf auf Grund seiner wissenschaftlichen Anforderungen und der seine Ausübung voraussetzenden Erfahrungen in der Patientenbetreuung für die Durchführung von Forschungen am Menschen qualifiziert. Wird eine Prüfung in einer Prüfstelle von mehreren Prüfern vorgenommen, so ist der verantwortliche Leiter der Gruppe der Hauptprüfer. Wird eine Prüfung in mehreren Prüfstellen durchgeführt, wird vom Sponsor ein Prüfer als Leiter der klinischen Prüfung benannt.

§ 26 Arzneimittelprüfrichtlinien

(1) Das Bundesministerium wird ermächtigt, nach Anhörung von Sachverständigen aus der medizinischen und pharmazeutischen Wissenschaft und Praxis durch Rechtsverordnung mit Zustimmung des Bundesrates Anforderungen an die in den §§ 22 bis 24, auch in Verbindung mit § 38 Abs. 2 bezeichneten Angaben, Unterlagen und Gutachten sowie

deren Prüfung durch die zuständige Bundesoberbehörde zu regeln. Die Vorschriften müssen dem jeweils gesicherten Stand der wissenschaftlichen Erkenntnisse entsprechen und sind laufend an diesen anzupassen, insbesondere sind Tierversuche durch andere Prüfverfahren zu ersetzen, wenn dies nach dem Stand der wissenschaftlichen Erkenntnisse im Hinblick auf den Prüfungszweck vertretbar ist. Die Rechtsverordnung ergeht, soweit es sich um radioaktive Arzneimittel und um Arzneimittel handelt, bei deren Herstellung ionisierende Strahlen verwendet werden und soweit es sich um Prüfungen zur Ökotoxizität handelt, im Einvernehmen mit dem Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit und, soweit es sich um Arzneimittel handelt, die zur Anwendung bei Tieren bestimmt sind, im Einvernehmen mit dem Bundesministerium für Ernährung, Landwirtschaft und Verbraucherschutz. Auf die Berufung der Sachverständigen findet § 25 Abs. 6 Satz 4 und 5 entsprechende Anwendung.

(2) Die zuständige Bundesoberbehörde und die Kommissionen nach § 25 Abs. 7 haben die Arzneimittelprüfrichtlinien sinngemäß auf das wissenschaftliche Erkenntnismaterial nach § 22 Abs. 3 und § 23 Abs. 3 Satz 2 anzuwenden, wobei die Besonderheiten der jeweiligen Arzneimittel zu berücksichtigen sind. Als wissenschaftliches Erkenntnismaterial gilt auch das nach wissenschaftlichen Methoden aufbereitete medizinische Erfahrungsmaterial.

§ 40 Allgemeine Voraussetzungen der klinischen Prüfung

(1) Der Sponsor, der Prüfer und alle weiteren an der klinischen Prüfung beteiligten Personen haben bei der Durchführung der klinischen Prüfung eines Arzneimittels bei Menschen die Anforderungen der guten klinischen Praxis nach Maßgabe des Artikels 1 Abs. 3 der Richtlinie 2001/20/EG einzuhalten. Die klinische Prüfung eines Arzneimittels bei Menschen darf vom Sponsor nur begonnen werden, wenn die zuständige Ethik-Kommission diese nach Maßgabe des § 42 Abs. 1 zustimmend bewertet und die zuständige Bundesoberbehörde diese nach Maßgabe des § 42 Abs. 2 genehmigt hat. Die klinische Prüfung eines Arzneimittels darf bei Menschen nur durchgeführt werden, wenn und solange

1. ein Sponsor oder ein Vertreter des Sponsors vorhanden ist, der seinen Sitz in einem Mitgliedstaat der Europäischen Union oder in einem anderen Vertragsstaat des Abkommens über den Europäischen Wirtschaftsraum hat,

2. die vorhersehbaren Risiken und Nachteile gegenüber dem Nutzen für die Person, bei der sie durchgeführt werden soll (betroffene Person), und der voraussichtlichen Bedeutung des Arzneimittels für die Heilkunde ärztlich vertretbar sind,
2. nach dem Stand der Wissenschaft im Verhältnis zum Zweck der klinischen Prüfung eines Arzneimittels, das aus einem gentechnisch veränderten Organismus oder einer Kombination von gentechnisch veränderten Organismen besteht oder solche enthält, unververtretbare schädliche Auswirkungen auf
 - a) die Gesundheit Dritter und
 - b) die Umweltnicht zu erwarten sind,
3. die betroffene Person
 - a) volljährig und in der Lage ist, Wesen, Bedeutung und Tragweite der klinischen Prüfung zu erkennen und ihren Willen hiernach auszurichten,
 - b) nach Absatz 2 Satz 1 aufgeklärt worden ist und schriftlich eingewilligt hat, soweit in Absatz 4 oder in § 41 nichts Abweichendes bestimmt ist und
 - c) nach Absatz 2a Satz 1 und 2 informiert worden ist und schriftlich eingewilligt hat; die Einwilligung muss sich ausdrücklich auch auf die Erhebung und Verarbeitung von Angaben über die Gesundheit beziehen,
4. die betroffene Person nicht auf gerichtliche oder behördliche Anordnung in einer Anstalt untergebracht ist,
5. sie in einer geeigneten Einrichtung von einem angemessen qualifizierten Prüfer verantwortlich durchgeführt wird und die Prüfung von einem Prüfer mit mindestens zweijähriger Erfahrung in der klinischen Prüfung von Arzneimitteln geleitet wird,
6. eine dem jeweiligen Stand der wissenschaftlichen Erkenntnisse entsprechende pharmakologisch-toxikologische Prüfung des Arzneimittels durchgeführt worden ist,
- 7.

jeder Prüfer durch einen für die pharmakologisch-toxikologische Prüfung verantwortlichen Wissenschaftler über deren Ergebnisse und die voraussichtlich mit der klinischen Prüfung verbundenen Risiken informiert worden ist,

8.

für den Fall, dass bei der Durchführung der klinischen Prüfung ein Mensch getötet oder der Körper oder die Gesundheit eines Menschen verletzt wird, eine Versicherung nach Maßgabe des Absatzes 3 besteht, die auch Leistungen gewährt, wenn kein anderer für den Schaden haftet, und

9.

für die medizinische Versorgung der betroffenen Person ein Arzt oder bei zahnmedizinischer Behandlung ein Zahnarzt verantwortlich ist.

Kann die betroffene Person nicht schreiben, so kann in Ausnahmefällen statt der in Satz 3 Nummer 3 Buchstabe b und c geforderten schriftlichen Einwilligung eine mündliche Einwilligung in Anwesenheit von mindestens einem Zeugen, der auch bei der Information der betroffenen Person einbezogen war, erteilt werden. Der Zeuge darf keine bei der Prüfstelle beschäftigte Person und kein Mitglied der Prüfgruppe sein. Die mündlich erteilte Einwilligung ist schriftlich zu dokumentieren, zu datieren und von dem Zeugen zu unterschreiben.

(2) Die betroffene Person ist durch einen Prüfer, der Arzt oder bei zahnmedizinischer Prüfung Zahnarzt ist, über Wesen, Bedeutung, Risiken und Tragweite der klinischen Prüfung sowie über ihr Recht aufzuklären, die Teilnahme an der klinischen Prüfung jederzeit zu beenden; ihr ist eine allgemein verständliche Aufklärungsunterlage auszuhändigen. Der betroffenen Person ist ferner Gelegenheit zu einem Beratungsgespräch mit einem Prüfer über die sonstigen Bedingungen der Durchführung der klinischen Prüfung zu geben. Eine nach Absatz 1 Satz 3 Nr. 3 Buchstabe b erklärte Einwilligung in die Teilnahme an einer klinischen Prüfung kann jederzeit gegenüber dem Prüfer schriftlich oder mündlich widerrufen werden, ohne dass der betroffenen Person dadurch Nachteile entstehen dürfen.

(2a) Die betroffene Person ist über Zweck und Umfang der Erhebung und Verwendung personenbezogener Daten, insbesondere von Gesundheitsdaten zu informieren. Sie ist insbesondere darüber zu informieren, dass

1.

die erhobenen Daten soweit erforderlich

a)

zur Einsichtnahme durch die Überwachungsbehörde oder Beauftragte des Sponsors zur Überprüfung der ordnungsgemäßen Durchführung der klinischen Prüfung bereitgehalten werden,

b)

pseudonymisiert an den Sponsor oder eine von diesem beauftragte Stelle zum Zwecke der wissenschaftlichen Auswertung weitergegeben werden,

c)

im Falle eines Antrags auf Zulassung pseudonymisiert an den Antragsteller und die für die Zulassung zuständige Behörde weitergegeben werden,

d)

im Falle unerwünschter Ereignisse des zu prüfenden Arzneimittels pseudonymisiert an den Sponsor und die zuständige Bundesoberbehörde sowie von dieser an die Europäische Datenbank weitergegeben werden,

2.

die Einwilligung nach Absatz 1 Satz 3 Nr. 3 Buchstabe c unwiderruflich ist,

3.

im Falle eines Widerrufs der nach Absatz 1 Satz 3 Nr. 3 Buchstabe b erklärten Einwilligung die gespeicherten Daten weiterhin verwendet werden dürfen, soweit dies erforderlich ist, um

a)

Wirkungen des zu prüfenden Arzneimittels festzustellen,

b)

sicherzustellen, dass schutzwürdige Interessen der betroffenen Person nicht beeinträchtigt werden,

c)

der Pflicht zur Vorlage vollständiger Zulassungsunterlagen zu genügen,

4.

die Daten bei den genannten Stellen für die auf Grund des § 42 Abs. 3 bestimmten Fristen gespeichert werden.

Im Falle eines Widerrufs der nach Absatz 1 Satz 3 Nr. 3 Buchstabe b erklärten Einwilligung haben die verantwortlichen Stellen unverzüglich zu prüfen, inwieweit die

gespeicherten Daten für die in Satz 2 Nr. 3 genannten Zwecke noch erforderlich sein können. Nicht mehr benötigte Daten sind unverzüglich zu löschen. Im Übrigen sind die erhobenen personenbezogenen Daten nach Ablauf der auf Grund des § 42 Abs. 3 bestimmten Fristen zu löschen, soweit nicht gesetzliche, satzungsmäßige oder vertragliche Aufbewahrungsfristen entgegenstehen.

(3) Die Versicherung nach Absatz 1 Satz 3 Nr. 8 muss zugunsten der von der klinischen Prüfung betroffenen Personen bei einem in einem Mitgliedstaat der Europäischen Union oder einem anderen Vertragsstaat des Abkommens über den Europäischen Wirtschaftsraum zum Geschäftsbetrieb zugelassenen Versicherer genommen werden. Ihr Umfang muss in einem angemessenen Verhältnis zu den mit der klinischen Prüfung verbundenen Risiken stehen und auf der Grundlage der Risikoabschätzung so festgelegt werden, dass für jeden Fall des Todes oder der dauernden Erwerbsunfähigkeit einer von der klinischen Prüfung betroffenen Person mindestens 500000 Euro zur Verfügung stehen. Soweit aus der Versicherung geleistet wird, erlischt ein Anspruch auf Schadensersatz.

(4) Auf eine klinische Prüfung bei Minderjährigen finden die Absätze 1 bis 3 mit folgender Maßgabe Anwendung:

1.

Das Arzneimittel muss zum Erkennen oder zum Verhüten von Krankheiten bei Minderjährigen bestimmt und die Anwendung des Arzneimittels nach den Erkenntnissen der medizinischen Wissenschaft angezeigt sein, um bei dem Minderjährigen Krankheiten zu erkennen oder ihn vor Krankheiten zu schützen. Angezeigt ist das Arzneimittel, wenn seine Anwendung bei dem Minderjährigen medizinisch indiziert ist.

2.

Die klinische Prüfung an Erwachsenen oder andere Forschungsmethoden dürfen nach den Erkenntnissen der medizinischen Wissenschaft keine ausreichenden Prüfergebnisse erwarten lassen.

3.

Die Einwilligung wird durch den gesetzlichen Vertreter abgegeben, nachdem er entsprechend Absatz 2 aufgeklärt worden ist. Sie muss dem mutmaßlichen Willen des Minderjährigen entsprechen, soweit ein solcher feststellbar ist. Der Minderjährige ist vor Beginn der klinischen Prüfung von einem im Umgang mit Minderjährigen erfahrenen Prüfer über die Prüfung, die Risiken und den Nutzen

aufzuklären, soweit dies im Hinblick auf sein Alter und seine geistige Reife möglich ist; erklärt der Minderjährige, nicht an der klinischen Prüfung teilnehmen zu wollen, oder bringt er dies in sonstiger Weise zum Ausdruck, so ist dies zu beachten. Ist der Minderjährige in der Lage, Wesen, Bedeutung und Tragweite der klinischen Prüfung zu erkennen und seinen Willen hiernach auszurichten, so ist auch seine Einwilligung erforderlich. Eine Gelegenheit zu einem Beratungsgespräch nach Absatz 2 Satz 2 ist neben dem gesetzlichen Vertreter auch dem Minderjährigen zu eröffnen.

4.

Die klinische Prüfung darf nur durchgeführt werden, wenn sie für die betroffene Person mit möglichst wenig Belastungen und anderen vorhersehbaren Risiken verbunden ist; sowohl der Belastungsgrad als auch die Risikoschwelle müssen im Prüfplan eigens definiert und vom Prüfer ständig überprüft werden.

5.

Vorteile mit Ausnahme einer angemessenen Entschädigung dürfen nicht gewährt werden.

(5) Der betroffenen Person, ihrem gesetzlichen Vertreter oder einem von ihr Bevollmächtigten steht eine zuständige Kontaktstelle zur Verfügung, bei der Informationen über alle Umstände, denen eine Bedeutung für die Durchführung einer klinischen Prüfung beizumessen ist, eingeholt werden können. Die Kontaktstelle ist bei der jeweils zuständigen Bundesoberbehörde einzurichten.

§ 41 Besondere Voraussetzungen der klinischen Prüfung

(1) Auf eine klinische Prüfung bei einer volljährigen Person, die an einer Krankheit leidet, zu deren Behandlung das zu prüfende Arzneimittel angewendet werden soll, findet § 40 Abs. 1 bis 3 mit folgender Maßgabe Anwendung:

1.

Die Anwendung des zu prüfenden Arzneimittels muss nach den Erkenntnissen der medizinischen Wissenschaft angezeigt sein, um das Leben dieser Person zu retten, ihre Gesundheit wiederherzustellen oder ihr Leiden zu erleichtern, oder

2.

sie muss für die Gruppe der Patienten, die an der gleichen Krankheit leiden wie diese Person, mit einem direkten Nutzen verbunden sein.

Kann die Einwilligung wegen einer Notfallsituation nicht eingeholt werden, so darf eine Behandlung, die ohne Aufschiebung erforderlich ist, um das Leben der betroffenen Person zu retten, ihre Gesundheit wiederherzustellen oder ihr Leiden zu erleichtern, umgehend erfolgen. Die Einwilligung zur weiteren Teilnahme ist einzuholen, sobald dies möglich und zumutbar ist.

(2) Auf eine klinische Prüfung bei einem Minderjährigen, der an einer Krankheit leidet, zu deren Behandlung das zu prüfende Arzneimittel angewendet werden soll, findet § 40 Abs. 1 bis 4 mit folgender Maßgabe Anwendung:

1.

Die Anwendung des zu prüfenden Arzneimittels muss nach den Erkenntnissen der medizinischen Wissenschaft angezeigt sein, um das Leben der betroffenen Person zu retten, ihre Gesundheit wiederherzustellen oder ihr Leiden zu erleichtern, oder

2.

a) die klinische Prüfung muss für die Gruppe der Patienten, die an der gleichen Krankheit leiden wie die betroffene Person, mit einem direkten Nutzen verbunden sein,

b) die Forschung muss für die Bestätigung von Daten, die bei klinischen Prüfungen an anderen Personen oder mittels anderer Forschungsmethoden gewonnen wurden, unbedingt erforderlich sein,

c) die Forschung muss sich auf einen klinischen Zustand beziehen, unter dem der betroffene Minderjährige leidet und

d) die Forschung darf für die betroffene Person nur mit einem minimalen Risiko und einer minimalen Belastung verbunden sein; die Forschung weist nur ein minimales Risiko auf, wenn nach Art und Umfang der Intervention zu erwarten ist, dass sie allenfalls zu einer sehr geringfügigen und vorübergehenden Beeinträchtigung der Gesundheit der betroffenen Person führen wird; sie weist eine minimale Belastung auf, wenn zu erwarten ist, dass die Unannehmlichkeiten für die betroffene Person allenfalls vorübergehend auftreten und sehr geringfügig sein werden.

Satz 1 Nr. 2 gilt nicht für Minderjährige, für die nach Erreichen der Volljährigkeit Absatz 3 Anwendung finden würde.

(3) Auf eine klinische Prüfung bei einer volljährigen Person, die nicht in der Lage ist, Wesen, Bedeutung und Tragweite der klinischen Prüfung zu erkennen und ihren Willen hiernach auszurichten und die an einer Krankheit leidet, zu deren Behandlung das zu prüfende Arzneimittel angewendet werden soll, findet § 40 Abs. 1 bis 3 mit folgender Maßgabe Anwendung:

1.

Die Anwendung des zu prüfenden Arzneimittels muss nach den Erkenntnissen der medizinischen Wissenschaft angezeigt sein, um das Leben der betroffenen Person zu retten, ihre Gesundheit wiederherzustellen oder ihr Leiden zu erleichtern; außerdem müssen sich derartige Forschungen unmittelbar auf einen lebensbedrohlichen oder sehr geschwächten klinischen Zustand beziehen, in dem sich die betroffene Person befindet, und die klinische Prüfung muss für die betroffene Person mit möglichst wenig Belastungen und anderen vorhersehbaren Risiken verbunden sein; sowohl der Belastungsgrad als auch die Risikoschwelle müssen im Prüfplan eigens definiert und vom Prüfer ständig überprüft werden. Die klinische Prüfung darf nur durchgeführt werden, wenn die begründete Erwartung besteht, dass der Nutzen der Anwendung des Prüfpräparates für die betroffene Person die Risiken überwiegt oder keine Risiken mit sich bringt.

2.

Die Einwilligung wird durch den gesetzlichen Vertreter oder Bevollmächtigten abgegeben, nachdem er entsprechend § 40 Abs. 2 aufgeklärt worden ist. § 40 Abs. 4 Nr. 3 Satz 2, 3 und 5 gilt entsprechend.

3.

Die Forschung muss für die Bestätigung von Daten, die bei klinischen Prüfungen an zur Einwilligung nach Aufklärung fähigen Personen oder mittels anderer Forschungsmethoden gewonnen wurden, unbedingt erforderlich sein. § 40 Abs. 4 Nr. 2 gilt entsprechend.

4.

Vorteile mit Ausnahme einer angemessenen Entschädigung dürfen nicht gewährt werden.

Auswertungsübersicht klinischer Studien 1933 – 1950

Gesamtstudienanzahl: 156			absolut	prozentual
Auswertungskriterien:				
Studientyp:		Fallserie	24	15
		Querschnittsstudie	7	4
		Outcomesstudie	60	38
		Experiment	67	42
		Andere	0	0
Kontrolle		Vorhanden	142	90
		Keine	16	10
		Historische	5	3
		Vorbeobachtung	6	4
		Nachbeobachtung	16	10
		Vor-+Nachbeobachtung	15	9
		Auslassversuch	3	2
		Verumvergleich	42	26
		Placebo	1	0,5
		Verum+Placebo	3	2
		Unbehandelt	45	28
		Anderes	6	4
Wurde ein Stichprobenplan beschrieben?	Ja	Explizit	14	9
		implizit	21	13
	Nein		123	78
Wurden Einschlusskriterien beschrieben?	Ja			
	explizit	Kriterien	31	20
		Keine Kriterien	0	0
	implizit	Kriterien	4	3
		Keine Kriterien	0	0
	Teilweise		32	20
	Nein		91	58
Wurden Ausschlusskriterien beschrieben ?	Ja			
	explizit	Kriterien	36	23
		Keine Kriterien	0	0
	implizit	Kriterien	5	3
		Keine Kriterien	0	0
	Teilweise		18	11
	Nein		99	63
Über welchen Zeitraum erstreckte sich die Studie?		Keine Angabe	60	38
		Tage	1	0,5
		Wochen	5	3
		Monate	54	34
		Jahre	35	22
		> 5 Jahre	3	2
Mehrrmige Studien			67	
Um welche Art der Gruppenbildung handelte es sich		Parallelgruppen	55	82
		Längsgruppen	4	6
		Keine Angabe	8	12
Wurde beschrieben, ob die Gruppen vergleichbar waren	Ja			
	explizit	Vergleichbar	16	24
		Nicht vergleichbar	8	5
	Implizit	Vergleichbar	24	36
		Nicht vergleichbar	0	
	Nein		0	

	Keine Angabe		19	28
Wurden die Gruppen gleichwertig behandelt?	Keine Angabe		20	30
	Ja			
	explizit	Gleichwertig	19	28
		Nicht gleichwertig	3	4
	implizit	Gleichwertig	25	37
		Nicht gleichwertig	0	0
	Nein		0	0
Wurde bei der Verteilung ein alternierendes Verfahren beschrieben?	Ja			
	explizit	Alternierung	12	18
	Implizit	Alternierung	11	16
	Nein		44	66
Wurde eine Verteilung nach bestimmten Gesichtspunkten beschrieben?	Ja			
	Explizit	Verteilung	29	43
		Keine Verteilung	0	0
	Implizit	Verteilung	3	4
		Keine Verteilung	0	0
	Nein		35	52
Einarmige Studien				
Wurden alle Patienten ansonsten gleichwertig behandelt?		Keine Angabe	39	
	Ja			
	Explizit	Gleichwertig	13	14
		Nicht gleichwertig	8	9
	implizit	Gleichwertig	23	25
		Nicht gleichwertig	4	4
	Nein		4	4
Wurden zusätzliche Medikamente gleichzeitig eingesetzt?		Keine Angaben	95	60
	Ja			
	explizit	Medikamente	45	28
		Keine Medikamente	8	5
	implizit	Medikamente	8	5
		Keine Medikamente	2	1
a) therapeutische beeinflussende Medikamente?		Keine Angaben	31	50
		Ja	20	32
		Nein	2	3
b) symptomatische Maßnahmen?		Keine Angabe	19	30
		Ja	32	50
		Nein	2	3
c) andere		Keine Angabe	46	73
		Ja	3	4
		Nein	4	6
Wurden die Outcomes operational definiert?	Ja		152	96
	Nein		6	4
	Keine Angabe		0	0
Zu welchem Zeitpunkt wurden die Outcomes bestimmt?		Keine Angabe	8	5
		Nach Studie	25	16
		Während Studie	9	6
		Beides	116	73
Wurde eine Verblindung der Behandlung beschrieben?	Ja			
	Explizit	Verblindung	5	3
		Keine Verblindung	0	
	Implizit	Verblindung	3	2
		Keine Verblindung	0	

Einarmige Studien:				
Wurden alle Patienten gleichwertig behandelt?	0 = keine Angabe	ja: 3 = nein	1 = explizit 1a = gleichw 1b=n.gleich	2 = implizit 2a=gleichfw 2b=n.gleich
7. Wurden weitere Medikamente zur selben Zeit eingesetzt?	0 = keine Angabe	ja: nein:	1 = explizit 3 = explizit	2 = implizit 4 = implizit
a) das Krankheitsbild im therapeutischen Sinne beeinflussende Medikamente	0 = keine Angabe	1 = ja	2 = nein	
b) symptomatische Maßnahmen	0 = keine Angabe	1 = ja	2 = nein	
c) andere	0 = keine Angabe	1 = ja	2 = nein	
8. Wurden die Outcomes operational definiert?	0 = keine Angabe	1 = ja	2 = nein	
9. Zu welchem Zeitpunkt wurden die Outcomes berstimmt?	0 = keine Angabe	1 = nach Ende der Studie		3 = 1 + 2
	2 = während der laufenden Studie			
10. Wurde eine Verblindung der Behandlung beschrieben?		ja: 3 = nein	1 = explizit	2 = implizit
Wurde eine Verblindung der Ergebnismessung und Auswertung beschrieben?		ja: 3 = nein	1 = explizit	2 = implizit
11. Wurde die Anwendung quantitativer/statistischer Auswertungsverfahren beschrieben?	0 = keine Angabe		1 = deskriptive Statistik 1a = einfach 2 = analytische Statistik 3 = 1 + 2 4 = nein	1b = komplex
12. Wurde über Subgruppenanalyse berichtet?		ja: 3 = nein	1 = explizit	2 = implizit
13. Wurde über eventuelle Dropouts berichtet?		ja: 3 = nein	1 = explizit 1a = aufgetr. 1b=n.aufgetr	2 = implizit 2a = aufgetr. 2b=n.aufgetr
Wurden in der Auswertung alle beteiligten Patienten berücksichtigt?	0 = keine Angabe	1 = ja	2 = nein	
Wurden die Patienten in der Auswertung ihrer Gruppe zugerechnet?	0 = keine Angabe	1 = ja	2 = nein	
14. Wurde über Nebenwirkungen berichtet?		ja: 2 = nein	1a = NW	1b = k. NW
15. Wurden die Ergebnisse kritisch diskutiert?		1 = ja	2 = nein	
16. Gedanken zu ethischen Problemen der Studiendurchführung?		1 = ja	2 = nein	
17. Gedanken zu methodologischen Aspekten?		1 = ja	2 = nein	
18. Wurden ausländische Autoren zitiert?			1 = angelsächsische	2 = andere 3 = keine

Danksagung

Die Idee, die zum Schreiben der vorliegenden Arbeit geführt hat, wurde durch Herrn Professor Raspe an mich herangetragen. Es bestand das Interesse, die Arbeit von Frau Dr. Stoll, die die klinische Forschung in Deutschland zwischen 1918 bis 1932 unter methodischen Aspekten untersucht hat, fortzuführen und Entwicklungen bis 1950 zu verfolgen.

Durch die Arbeit hatte ich die einmalige Gelegenheit, ein tieferes Verständnis für den historischen Entwicklungsprozess klinischer Forschung zu entwickeln.

Zunächst gilt mein Dank all den Personen, die das Zustandekommen der Arbeit durch ihre individuelle Hilfsbereitschaft möglich gemacht haben. Insbesondere möchte ich Herrn Professor Roelcke danken, der mir während eines Besuchs am Institut für Geschichte der Medizin an der Universität in Giessen zu Beginn der Arbeit wertvolle Hinweise im Umgang mit den Quellen gegeben hat. Ferner gilt mein Dank Herrn Professor von Engelhardt für seine logistische Hilfe.

Ganz wesentliche Unterstützung habe ich durch Herrn Professor Borck und durch die Mitarbeiterinnen und Mitarbeiter des Instituts für Medizingeschichte und Wissenschaftsforschung der Universität Lübeck erfahren. Dankbar bin ich insbesondere für die anregenden Gespräche mit Herrn Professor Borck, aus denen ich viele wertvolle Hinweise und Informationen, die wesentlich zum Verständnis des historischen Kontexts beigetragen haben, gewinnen konnte. Den Mitarbeitern des Instituts und ganz besonders Frau Hoffmann und Frau Österreich möchte ich für ihre Hilfsbereitschaft danken.

Frau Dr. Stoll gilt ein besonderer Dank. Zum einen hat sie mir zu Beginn der Arbeit wertvolle Hilfestellung geleistet, indem sie mir wichtige Hinweise über die von ihr angewandte Methodik gegeben hat, die für die vorliegende Arbeit maßgeblich ist. Zum anderen hat sie durch ihre Arbeit mein Interesse an der geschichtlichen Entwicklung klinischer Versuche geweckt.

Den Mitarbeiterinnen und Mitarbeitern des Instituts für Sozialmedizin bin ich für ihre hilfsbereite und freundliche Art zu großem Dank verpflichtet. Dadurch war es mir möglich, diese Arbeit in einer sehr guten und kollegialen Atmosphäre anzufertigen.

Mein herzlichster Dank gilt Herrn Professor Raspe, der diese Arbeit auf sehr geduldige und verständnisvolle Art begleitet und unterstützt hat. In vielen konstruktiven Gesprächen konnten Probleme offen diskutiert und gelöst werden. So bin ich in ganz besonderem Maße Herrn Professor Raspe dankbar, dass die Arbeit trotz anfänglicher Schwierigkeiten zum Abschluss gebracht werden konnte.

Lebenslauf

Persönliche Daten

Name: Christoph Wolkewitz
Geburtsdatum: 23.08.1981
Geburtsort: Eschwege
Wohnort: Bergstraße 5
38102 Braunschweig



Schulbildung

1988 – 1992 Besuch der Grundschule in Celle
1992 – 2001 Besuch des Gymnasium Ernestinum in Celle
2001 Abitur

Studium

10/2001 – 07/2002 Studium der Volkswirtschaftslehre in Regensburg
10/2002 – 11/2008 Studium der Humanmedizin in Lübeck

Examina

08/2004 Ärztliche Vorprüfung
11/2008 Ärztliche Prüfung

Praktisches Jahr

- 08/2007 – 12/2007 1. Tertial: Innere Medizin: Rheumatologie in Bad Bramstedt
- 12/2007 – 04/2008 2. Tertial: Chirurgie: Unfallchirurgie/Viszeralchirurgie
Universitätsklinikum Schleswig –Holstein Campus Lübeck
- 04/2008 – 07/2008 3. Tertial: Orthopädie: Universitätsklinikum Schleswig – Holstein
Campus Lübeck

Dissertation

- Seit 06/2005 Beginn der Arbeit an der vorliegenden Dissertation zum Thema
„Die Methodik klinischer Forschung im frühen 20. Jahrhundert“

Berufliche Tätigkeit

- 12/2008 – 04/2009 Wissenschaftliche Hilfskraft am Institut für Sozialmedizin der
Universität Lübeck
- Seit 04/2009 Assistenzarzt in der Allgemein- und Viszeralchirurgie im Klinikum
Peine in der Weiterbildung zum Allgemeinmediziner

Sprachkenntnisse: Deutsch (Muttersprache), Englisch