

Evidenzbasierung in der medizinischen Rehabilitation: eine systematische Literaturübersicht am Beispiel der Indikation chronischer Rückenschmerz

Bericht an den Sachverständigenrat zur Begutachtung der Entwicklung im Gesundheitswesen (im Bundesministerium für Gesundheit)

Heiner Raspe und Angelika Hüppe, Universität zu Lübeck

Kapitel 1: Hintergrund und Ziele des Forschungsvorhabens

Medizinische Rehabilitation ist eine Grundfunktion, eine basale „Gesundheitsstrategie“ (Meyer et al. 2011) jedes entwickelten Gesundheitssystems. Der World Report on Disability der WHO und Weltbank (WHO 2011) versteht unter Rehabilitation ein „Bündel von Maßnahmen, das Individuen, die Behinderungen haben oder wahrscheinlich haben werden, dabei unterstützt, ein optimales Funktionsniveau im Austausch mit ihrer Umgebung zu erreichen und zu erhalten“ [1].

In Deutschland ist dieser Funktionsbereich, sofern er in der Verantwortung verschiedener Zweige unseres Sozialversicherungssystems steht, in SGB IX geordnet. Wesentliche Träger der Rehabilitation sind (neben der Gesetzlichen Unfallversicherung) die Deutsche Renten- und die Gesetzliche Krankenversicherung (DRV, GKV). Sie sind jeweils Strukturverantwortliche und Kostenträger und in geringem Umfang auch Leistungserbringer in einem. Der größte Teil der von der DRV gewährten regelmäßig multidisziplinären und multimodalen Leistungen bezieht sich auf Versicherte mit muskuloskelettalen Erkrankungen (31 %) und hier vor allem auf unspezifische Rückenschmerzen (ICD M50-54: 54 %, d.h. 17 % vom Gesamt) (Raspe 2012).

Hier steht die Rehabilitationsform „Heilverfahren“ in Form eines im Mittel dreiwöchigen Aufenthalts in einer der weit mehr als 500 wohnortfernen stationären Einrichtungen, unter vorzugsweise orthopädischer Führung, ganz im Vordergrund. Dies gilt auch für diese Literaturübersicht. Allerdings sind Studien zur Anschlussheilbehandlung nicht grundsätzlich ausgeschlossen. Beide sind komplexe Interventionen im Sinne der Empfehlungen des englischen Medical Research Councils (MRC 2008).

¹ „a set of measures that assist individuals who experience, or are likely to experience, disability to achieve and maintain optimal functioning in interaction with their environments“.

Das Konzept der evidenzbasierten Medizin (EbM) wurde in Deutschland erstmals 1995/96 rezipiert (Klemperer 1995; Raspe 1996). Es fordert von jedem Funktionsbereich der Medizin zwischen Prävention und Rehabilitation, dass seine Leistungen „dem allgemein anerkannten Stand der medizinischen Erkenntnisse“ (§ 2 Abs. 1 Satz 3 SGB V) entsprechen. Als dessen zentrales Fundament gelten heute die Ergebnisse hochwertiger und öffentlich zugänglicher klinischer und Versorgungsstudien. Ihr Ziel ist eine möglichst unverzerrte Schätzung des wahren mittleren (differentiellen) Effekts einzelner Untersuchungs- und Behandlungsmethoden und komplexer Interventionen. Die in ihnen enthaltenen Informationen und Daten werden in einem aufwändigen und kritischen (inter)nationalen Diskussionsprozess zu dem, was aktuell als „Evidenz“ bezeichnet wird. Mit solcher Evidenz lässt sich die Richtigkeit von Behauptungen – etwa zur Wirksamkeit und zum Nutzen einer medizinisch-beruflichen Rehabilitation in Trägerschaft der DRV – vor einem kritischen Publikum mehr oder weniger überzeugend verteidigen.

Dazu bedarf es Studien, die grundsätzlich dreierlei leisten müssen: einerseits müssen sie eine Ursache-Wirkungsbeziehung („generelle Wirksamkeit“) zwischen rehabilitativer Intervention und empirisch erfassten Ergebnissen belegen; zweitens müssen sie das Ausmaß des erzielten (Zusatz)Nutzens möglichst genau abschätzen lassen. Und drittens müssen sie in mehrfacher Hinsicht vertrauenswürdig sein.

Unsere Literaturübersicht hat das Ziel, die methodische Qualität der in Deutschland durchgeführten klinischen und Versorgungs-Studien zur Rehabilitation in den Jahren 2000 bis 2013 darzustellen und zu bewerten. Dabei sollen vereinbarungsgemäß nur solche interventionelle Studien eingeschlossen werden, die – mit oder ohne ökonomische (Zusatz)-Untersuchungen – eine Kontrollgruppe und Kontrollbedingung mitführten.

Im Einzelnen sind folgende vier Forschungsfragen zu behandeln:

1. Wie ist der Stand der Evidenzbasierung von medizinischen Reha-Maßnahmen in Deutschland (am Beispiel der Indikation chronischer Rückenschmerz)?
2. Entspricht der methodische Standard der vorliegenden Studien den Ansprüchen, die man für ein HTA (zur medizinischen Rehabilitation) anlegen würde?
3. Gibt es besondere Lücken oder Schwächen in der Studienlage aus dem deutschen Gesundheitssystem?
4. Welche Empfehlungen können aus der ermittelten Studienlage hinsichtlich eines idealen methodischen Designs zukünftiger Evaluationen abgeleitet werden?

Das Vorhaben setzt auf Vorarbeiten von Hüppe und Raspe aus den Jahren 2003 und 2005 auf (Hüppe & Raspe 2003, 2005). Damals wurden insgesamt 46 relevante Studien (Zeitraum Januar 1980 bis Dezember 2003) identifiziert und ausgewertet (unter ihnen befanden sich 22 unkontrollierte Studien). Nur in einer wurde die aktuelle stationäre Rehabilitation mit der Bedingung „keine Rehabilitation“ im Rahmen eines kurzfristigen Wartegruppendedesigns verglichen (Jäckel et al. 1990). Während von uns damals ausschließlich stationäre Formen der Rehabilitation einbezogen wurden, werden in der aktuellen Arbeit auch ambulante bzw. teilstationäre Angebote medizinischer Rehabilitation berücksichtigt.

Von diesen Vorarbeiten unterscheidet sich das folgende Gutachten in einer wesentlichen Hinsicht: Es kommentiert nicht die in den einzelnen Studien gefundenen Ergebnisse, sondern konzentriert sich – vertragsgemäß – ausschließlich auf die Methodik der Studien und deren Qualität. Dabei geht es in der evidenzbasierten Medizin und auch in HTAs natürlich nicht um eine allgemeine Beurteilung der methodischen Qualität der Studien zu einem bestimmten Versorgungs- und Forschungsfeld, sondern um die verlässliche Beantwortung gezielter klinischer oder versorgungspolitischer Fragen. Und hierfür spielt die Präzision und Glaubwürdigkeit der Studienergebnisse („Ergebnissicherheit“) eine hervorragende Rolle.

Dabei gilt, dass die methodischen Ansprüche an die Studien mit abnehmender Effektstärke zunehmen: „Je kleiner der zu erwartende Effekt einer Intervention jedoch ist, desto höher werden die Anforderungen an die Ergebnissicherheit sein ...“ (Lange & Thomas 2010). Nach allem, was wir über die Wirksamkeit und den Nutzen der medizinischen Rehabilitation bei chronischen Rückenschmerzen wissen, ist bisher nur mit durchschnittlich schwachen bis moderaten Effekten zu rechnen.

Kapitel 2 Material und Methoden

2.1 Literatursuche und -auswahl

2.1.1 Kriterien für den Studieneinschluss:

Es wurden nur Primärstudien eingeschlossen, zu denen Ergebnisse in einem wissenschaftlichen Fachjournal im Zeitraum 2000 bis 10/2013 in deutscher oder englischer Sprache publiziert wurden und die Aussagen zu differentiellen Effekten, Wirksamkeit, (Netto)Nutzen oder Effizienz von Rehabilitationskonzepten oder -modulen treffen. Dissertationen sowie Abschlussberichte wurden berücksichtigt, sofern sie ergänzende Informationen zu einem publizierten Manuskript liefern konnten.

Das Review berücksichtigt nur interventionelle und vergleichende Studien. In Anlehnung an § 4 Abs. 23 AMG verstehen wir darunter wenigstens zweiarmig-vergleichende Untersuchungen, in denen die Behandlung einschließlich der Diagnose und Überwachung eines vorab definierten Teils der Patienten einem von den Wissenschaftlern oder ihren Auftraggebern festgelegten Prüfplan folgt und nicht der eingespielten Routine (wie in einer „Anwendungsbeobachtung“) überlassen wird.

Damit kommen sowohl randomisierte kontrollierte Studien (RCT) wie auch nicht-randomisierte vergleichende Studien in Frage. Zu dieser zweiten Gruppe zählten wir Untersuchungen, die eines der folgenden „quasiexperimentellen“ Designs (Shadish et al. 2002) aufwiesen: sequentielles Kontrollgruppendesign (Interventionsgruppe (IG) und Kontrollgruppe (KG) zeitlich versetzt) oder paralleles Kontrollgruppendesign (IG und KG zeitlich parallel) jeweils mit Vorher- und Nachher-Messungen der Ergebnisparameter.

Wenigstens ein Studienarm muss Teilnehmende an einer in Deutschland stationär oder ambulant durchgeführten multimodalen medizinischen Rehabilitationsmaßnahme umfassen, die als Anschlussheilbehandlung oder Heilverfahren nach den speziellen gesetzlichen (Zugangs)Voraussetzungen gewährt und durchgeführt wurde. Alle Studienteilnehmende (oder wenigstens der überwiegende Teil) sollten wegen „sonstiger Krankheiten der Wirbelsäule und des Rückens“ (ICD 10: M50-54) oder einer anderen mit Rückenschmerzen verbundenen muskuloskelettalen Erkrankung behandelt worden sein.

2.1.2 Suchstrategie

Zur Anwendung kamen elektronische Literaturrecherche und Handsuche.

Elektronische Recherche:

Über DIMDI smartsearch wurde Anfang November 2013 auf die Datenbanken MEDLINE, PsychINFO, PSYINDEX und EMBASE zugegriffen.

Die Schlagworte für die Recherche waren fünf verschiedenen Modulen zugeordnet (vgl. Tabelle 1). Die Recherche folgte zwei verschiedenen Verknüpfungen der fünf Module: Verknüpfung 1 = (A +B +C) + E sowie Verknüpfung 2 = (A + D) + E.

Die Schlagworte innerhalb eines Moduls wurden jeweils mit „OR“ verknüpft, die Module untereinander mit „AND“. Im Modul A („Maßnahme“) und E („Eingrenzung“) wurden neben einer Suche mit Freitexten (FT) auch Thesaurussuchen (CT bzw. CTG) genutzt.

Um nur Studien zu berücksichtigen, die im Rahmen des deutschen Gesundheitssystems durchgeführt wurden, erfolgte eine Eingrenzung der Suchergebnisse über die Verknüpfung mit den Schlagworten: ... AND (deutsch* OR german* OR LANGUAGE: german).

Tabelle 1: Schlagworte der elektronischen Literaturrecherche

Modul A Maßnahme	Modul B Wirksamkeit	Modul C Studiendesign	Modul D ökonomisch	Modul E Eingrenzung
reha*	wirksam* nutzen* benefit* utility* effective* effectivity* efficacy* efficien* effizien* effektiv* Bewertung Evaluation	control* kontrol* randomi* RCT CCT Meta-Analyse meta-analysis Metaanalyse metaanalysis	cost* kosten* econom* ökonom* oekonom* gesundheitsökonom* gesundheitsoekonom*	Rückenschmerz* Kreuzschmerz* Rückenleiden back pain dorsal pain spine* dorsalgia

Handsuche:

Folgende Quellen wurden nach Hinweisen auf einschlägige Studien durchgesehen:

- Tagungsbände der rehabilitations-wissenschaftlichen Kolloquien (2000 – 2013),
- Systematische Reviews in der Cochrane-Review-Datenbank, die einen Bezug zur Rehabilitation aufweisen und im Titel/Abstract einen Hinweis auf den Bereich Rücken oder Wirbelsäule enthalten,
- Literaturübersichten zur Wirksamkeit bzw. Effizienz medizinischer Rehabilitationsmaßnahmen aus der oben genannten elektronischen Literaturrecherche.

Auf eine Durchsicht der Bände der Zeitschrift „Rehabilitation“ wurde verzichtet, da alle in ihr veröffentlichten Originalarbeiten in MEDLINE gelistet sind.

Wir danken an dieser Stelle Herrn Dr. Hodek, ehemals Mitarbeiter der Geschäftsstelle des Sachverständigenrates zur Begutachtung der Entwicklung im Gesundheitswesen (BMG) für seine vielfältige Unterstützung bei der Literatursuche und -beschaffung: Herr Hodek führte die zuvor gemeinsam mit HR und AH abgestimmten Datenbanksuchen durch; er sichtete die Tagungsbände der Rehakolloquien und stellte zur weiteren Bearbeitung die Abstract-Texte zusammen, die einen Hinweis auf die Suchbegriffe "Reha" und "Rückenschmerz" aufwiesen. Er recherchierte in der Cochrane Library nach einschlägigen Reviews. Zu den von HR und AH ausgewählten Literaturstellen beschaffte er Kopien für die weitere Bearbeitung. Wir danken weiterhin Frau Dr. Lühmann, Mitarbeiterin des Instituts für Allgemeinmedizin am Universitätskrankenhaus Hamburg-Eppendorf für ihre Beratung bei der Entwicklung der Suchstrategie.

Studienselektion

In einem ersten Schritt wurden alle Treffer der elektronischen Recherche anhand der Abstracts bzw. der Titel (falls kein Abstract vorhanden) auf Mehrfachnennung geprüft, und es wurden Dubletten ausgeschlossen. Aus dem Kreis der verbliebenen Referenzen wurden solche aussortiert,

- die ausschließlich von Studien berichten, die im Rahmen der ambulanten Versorgung (z.B. in Schmerzambulanzen), also außerhalb eines Antragsverfahrens bei den Rehabilitationsträgern, durchgeführt wurden,
- die keine Kontroll- oder Vergleichsgruppe mitführten oder
- keinen Bezug zum Thema Rehabilitation erkennen ließen.

Trafen die Kriterien für einen Studieneinschluss auf der Basis der zur Verfügung stehenden Informationen zu, wurden Volltexte beschafft. Für die Handsuche wurden zusätzlich thematisch einschlägige Literaturübersichten bestellt.

In einem zweiten Auswahlschritt wurden die Volltexte an den unter 2.1.1 genannten Studieneinschlusskriterien überprüft. Ebenso wurden Studien, die alle Einschlusskriterien erfüllten, eine Identifikationsnummer (IDN) zugewiesen. Lagen zu einer Studie mehrere Publikationen vor, wurde diesen Publikationen dieselbe IDN zugeordnet.

2.2 Studienbewertung

Der Auftrag stellt die Frage, ob „der methodische Standard der vorliegenden Studien den Ansprüchen (entspricht), die man für ein HTA anlegen würde“ – wir interpretieren: für ein HTA zu Wirksamkeit, (Netto)Nutzen und ggf. Effizienz der medizinisch-beruflichen Rehabilitation bei chronischen Rückenschmerzen. Nun ist die Klärung der entsprechenden Evidenzlage zwar ein zentrales, aber eben auch nur ein Kapitel eines HTAs, und hierzu gibt es keinen eigenen Goldstandard (siehe (Dreier et al. 2010; Bai et al. 2012)).

2.2.1 Interne Validität (methodische Studienqualität)

Zur Bewertung der internen Validität der einbezogenen Studien folgten wir den 2009 überarbeiteten Vorschlägen der Cochrane Back Review Group CBRG (Furlan et al. 2009). Zum Einsatz kam eine Liste mit zwölf von der CBRG operational definierten Kriterien (s. Furlan et al 2009, table 3), die sich am Cochrane Risk of Bias Tool des aktuellen Cochrane Handbuchs für systematische Reviews von Interventionsstudien orientieren (Higgins & Green 2011). Jedes Kriterium wurde mit den drei Antwortkategorien ja (=erfüllt), nein (=nicht erfüllt) oder unklar (=unzureichende Informationen) bewertet. Es wird davon ausgegangen, dass mit der Anzahl erfüllter Kriterien die Vertrauenswürdigkeit der Studienergebnisse steigt, der Studie wird eine höhere interne Validität zugeschrieben; andersherum: je weniger Kriterien erfüllt sind, umso höher ist die Wahrscheinlichkeit systematisch verfälschter Ergebnisse. Eine geringe interne Validität geht also einher mit einem hohen Verzerrungsrisiko (risk of bias).

Nach den Vorgaben von Furlan et al (2009) sprechen wir einer Studie eine hohe interne Validität (und damit ein vergleichsweise geringes Verzerrungsrisiko) zu, wenn mindestens sechs der zwölf Kriterien als erfüllt bewertet werden. Bei einer Studie mit weniger als sechs mit „Ja“ bewerteten Kriterien wird hingegen ein erhöhtes Bias-Risiko vermutet. Die Bewertung der einzelnen Kriterien der internen Validität wird für jede Studie transparent dargestellt.

2.2.2 Weitere Qualitätsaspekte

Zusätzlich zur Bewertung der internen Validität wurden weitere Aspekte erfasst, wie sie z.T. auch zur Einschätzung der Berichtsqualität einer Publikation herangezogen werden (vgl. z.B. Checklisten CONSORT 2010 (Moher et al. 2010); STROBE (von Elm et al. 2007)). Eine adäquate Berichtsqualität gilt als Voraussetzung für eine belastbare Bewertung der methodischen Studienqualität.

Dokumentiert wurden: Studienziel? Untersuchte Rehabilitationsformen bzw. -Leistungen/ Module? Art des Studiendesigns? Mono- oder multizentrisch durchgeführt? Ein- und Ausschlusskriterien klar definiert, eng oder weit gefasst? Fallzahl begründet? Intervention präzise beschrieben? Vorliegen des Votums einer Forschungsethikkommission? Start der Rekrutierung? Sozialmedizinisch relevante Outcome-Parameter (z.B. soziale Teilhabe)? Dauer des Follow-Up (ausreichend langer Katamensezeitraum, mindestens zwölf Monate)? Charakteristika der Studienteilnehmenden (Alter, Geschlecht, Diagnosen)? Kohortenverluste? Potenzieller Schaden angesprochen (negative Behandlungseffekte)? Gesundheitsökonomische (Zusatz)Analysen? Diskussion methodischer Limitationen der Studie?

2.3 Datenextraktion

Zu jeder eingeschlossenen Einzelstudie wurden die oben genannten Informationen exzerpiert und auf einem Dokumentationsblatt notiert. Bei unklaren oder fehlenden Informationen wurde – falls vorhanden – auf öffentlich zugängliche Abschlussberichte zum Projekt zurückgegriffen. Nur in Einzelfällen wurden die für die Studien verantwortlichen Wissenschaftler kontaktiert, um Details zu klären.

Datenextraktion und Qualitätsbewertung wurden jeweils von zwei unabhängigen Personen (HR und AH) vorgenommen, bei Diskrepanzen erfolgte eine Konsensentscheidung.

Zur Bewertung möglicher Entwicklungen über den Beobachtungszeitraum dieses Reviews wurden die Studien nach dem Jahr des Rekrutierungsbeginns in drei Gruppen geteilt: vor 2000, 2000-2005, nach 2005.

Die erfassten Daten und Beurteilungen wurden unter Anwendung des Statistikpaketes IBM SPSS Version 22 ausgewertet.

Kapitel 3 - Ergebnisdarstellung

3.1 Suchergebnisse

Die elektronische Literaturrecherche führte zu insgesamt 540 Treffern (hits), nach Abzug mehrfach identifizierter Referenzen verblieben 399 Treffer (vgl. Tabelle 2).

Tabelle 2: Ergebnis der elektronischen Literaturrecherche (05.11.2013)

Anzahl Treffer: Verknüpfung (A + B + C) + E	Anzahl Treffer: Verknüpfung (A + D) + E
6.777 Treffer nach Eingrenzung auf deutsche Studien 350 Treffer davon Embase: 139 davon Medline: 167 davon PsycINFO: 8 davon Psyn dex: 36	2.650 Treffer nach Eingrenzung auf deutsche Studien 190 Treffer davon Embase: 116 davon Medline: 60 davon PsycINFO: 1 davon Psyn dex: 13
Um Dubletten bereinigt 399 Treffer	

Ergänzt um die Ergebnisse der Handsuche lagen nach dem ersten Auswahl schritt (Prüfung der Ein-/Ausschlusskriterien anhand der Abstract-Texte) insgesamt 68 Volltexte zur weiteren Bearbeitung vor. Der zweite Auswahl schritt identifizierte 28 verschiedene Studien, zu denen in 50 Texten (darunter zwölf englischsprachige), berichtet wurde. Eine Aufstellung dieser 28 Studien mit den zugehörigen Publikationen und den ihnen zugeordneten IDN findet sich in Tabelle A1 im Anhang.

Tabelle A2 im Anhang führt die von der weiteren Bearbeitung ausgeschlossenen Texte auf, die sich aus 14 Literaturzusammenfassungen, einer unkontrollierten Studie, einer nicht-interventionellen Studie, einer Studie aus Österreich sowie einer Dissertation ohne zugehöriger Zeitschriftenpublikation zusammensetzen.

3.2 Steckbrief der eingeschlossenen Studien

Die 28 eingeschlossenen Studienprojekte begannen mit der Rekrutierung der Studienteilnehmenden zwischen 1995 und 2009; sie verteilen sich wie folgt auf die drei genannten Zeiträume: elf begannen vor 2000, neun zwischen 2000 – 2005 und acht nach 2005.

3.2.1 Studiengruppen und Studienzentren

1998 nahmen die vom BMBF und der DRV geförderten Rehabilitationsforschungsverbände ihre Arbeit auf. Nur fünf der eingeschlossenen Studien lassen sich keinem der acht Verbände oder ihrem Umfeld zuordnen. Es sind dies vor allem die Studien, die der Standardrehabilitation ein isoliertes Element (z.B. Akupunktur) hinzugefügt hatten. Am häufigsten sind die Verbände aus Westfalen und Südwestdeutschland mit je fünf Studien vertreten.

Die Zahl der in die Studien involvierten Zentren liegt zwischen eins und sechs. 13 Studien waren allein monozentrisch organisiert.

3.2.2 Studiendesigns

Projekte der Rehabilitationsforschung weisen Besonderheiten in ihrem Design auf. Diese erschweren es, sie etablierten Taxonomien, etwa der der Cochrane Bias Methods Group, zuzuordnen.

Auftragsgemäß haben wir nur interventionelle Studien nach o.g. Definition aufgenommen. Wir unterscheiden vier Typen: 1. randomisierte kontrollierte Studien (Randomisierung von einzelnen Personen oder Personen-„Clustern“), 2. Studien mit sequentieller/n Kontrollgruppe/n, 3. Studien mit paralleler Kontrollgruppe und 4. Studien mit paralleler Kontrastgruppe (vgl. Tabelle 3).

Das zweite Design ist eines, das uns in der Rehabilitationsforschung erstaunlich oft begegnet – sonst aber in der klinischen Forschung selten genutzt wird: Interventions- und Kontrollgruppe einer Quelle (z.B. Versicherte eines Rehaträgers) werden in einer oder mehreren Rehakliniken nacheinander rekrutiert und für definierte Zeitperioden und in definierten Zeitabständen einer von der bisherigen Routine abweichenden Intervention exponiert. Im einfachsten Fall bleibt es bei einem Wechsel (Kontrollgruppe vor Interventionsgruppe); jedoch sind mehrere durchaus möglich.

Der – graduelle - Unterschied zwischen dem Typus 3 (parallele Kontrollgruppe) und Typus 4 (parallele Kontrastgruppe) liegt darin, dass wir bei den Studien des 3. Typus grundsätzlich von einer Ähnlichkeit der parallel beobachteten Gruppen ausgehen können, während die Teilnehmenden der Kontrastgruppenstudien (Typ 4) definitiv aus einer unterschiedlichen Population stammen. Damit nimmt das Biasrisiko für direkte Gruppenvergleiche von Typ 3 zu Typ 4 noch deutlich zu. Im Hintergrund steht die in der Methodology Checklist 3 des Scottish Intercollegiate Guidelines Networks formulierte Frage 1.2: „(Sind) die beiden Studiengruppen aus Quellpopulationen rekrutiert, die in allen Aspekten außer dem zu untersuchenden Faktor

vergleichbar sind?“ [2] (SIGN 2013).

Von den 28 eingeschlossenen Studien folgten zwölf einem kontrollierten-randomisierten Design (vgl. Tabelle 3).

Tabelle 3: Zuordnung der Studien (IDN) zu Studiendesign und Projektstart

Studiendesign		Projektstart vor 2000	Projektstart 2000-2005	Projektstart nach 2005
randomisierte kontrollierte Studie	(N=12)	06,13,14,16,20,22	12,23,25,28	18,26
sequentielles Kontrollgruppendesign	(N=10)	15,24	01,05,10,17	07,08,09,27
paralleles Kontrollgruppendesign	(N= 2)			04,21
paralleles Kontrastgruppendesign	(N= 4)	02,03,19	11	

Die Besonderheit der einzigen „Cluster-randomisierten“ Studie (IDN 13) ist, dass feste Gruppen von (einer einzigen Klinik konsekutiv zugewiesenen) Patienten in dieser - randomisiert - nacheinander in vielfachem Wechsel der Verum- oder der Kontrollbedingung zugewiesen wurden. Von den 16 nicht randomisierten Studien lassen sich zehn dem sequentiellen Kontrollgruppen-Design zuordnen; davon wechselt die Hälfte nur einmal zwischen Kontroll- und Interventionsgruppe. Sechs Studien führen die Intervention in zuvor bestimmten Einrichtungen durch und erheben zeitlich parallel Daten in Vergleichseinrichtungen, die die Intervention nicht durchführen. Vier von ihnen lassen klar erkennen, dass nicht von einer Kontrollgruppe, sondern höchstens von einer Kontrastgruppe gesprochen werden kann (Typ 4).

Zwei Studien (IDN 03,06) waren vierarmig angelegt, vier dreiarig (IDN 09,11,20,23), der Rest zweiarig.

3.2.3 Fragestellungen und Ziele der Studien

Am häufigsten (22 von 28) beobachteten wir den Vergleich der üblichen stationären oder ambulanten „Standardrehabilitation“ mit einer „angereicherten“ Rehaform (Reha A vs. Reha A plus x). Dabei wird zehnmals eine stationäre Rehabilitation „nach den Vorschriften des Rehaträgers“ mit einem umfassend veränderten ebenfalls stationären Rehabilitationsprogramm (das aber weiter den Vorschriften entspricht und damit erhebliche Ähnlichkeiten zum Standard aufweist) verglichen (z.B. IDN 05,13); siebenmal wird die stationäre, einmal die ambulante Standardrehabilitation um einige wenige, gut abgrenzbare Leistungen bzw. Module ergänzt: dabei handelt es sich um Akupunktur (IDN 20,26), computerbasiertes Selbstregulationsprogramm (IDN 21), kognitiv-behaviorale Behandlung (IDN 16), Depres-

² “(Are) the two groups being studied ... selected from source populations that are comparable in all respects other than the factor of investigation.”

sionsbewältigungstraining (IDN 9), muskuläre Stabilisierung (IDN 12), Rückenschule (IDN 18), Akupunktmassage (IDN 06). Eine weitere Studie prüft eine ausdrückliche Zielvereinbarung zu Rehabeginn (IDN 07), zwei Studien konzentrieren sich auf die Prüfung unterschiedlicher Zuweisungsmodalitäten (IDN 22,23) und eine evaluiert ein komplexes Nachsorgeprogramm (IDN 04).

Viermal wird die klassische stationäre mit einer teilstationären oder ambulanten Rehabilitation verglichen (IDN 02,03,14,19). Eine Studie vergleicht verschiedene ambulante Programme (IDN 11). Eine Studie (IDN 28) stellt die rehabilitative Versorgung der Versorgung ohne Rehabilitation (usual care) gegenüber.

Insgesamt fanden 21 Studien ausschließlich im stationären Rehasetting statt, sechs Studien berücksichtigten (auch) eine teilstationäre/ambulante Rehaform (IDN 02,03,11,12,14,19), eine Studie unterschied nicht zwischen stationärer oder ambulanter Rehaform (IDN 28).

3.2.4 „Absolute“ und „relative“ Wirksamkeit

Wirksamkeit kann – von Grenzfällen abgesehen - nur vergleichend bestimmt werden. Eine Wirksamkeitsstudie muss die Frage beantworten können, was sich ergibt (und in Zukunft zu erwarten ist), wenn eine möglichst identische Gruppe von Patienten gar nicht oder nur scheinbar behandelt wird. Dieser kontrafaktische Ansatz ist im Bereich der Arzneimittelstudien, so weit es geht, im Placebo-kontrollierten RCT mit mehrfacher Verblindung realisiert.

Dieses Design ist im Bereich der Rehabilitationsforschung nur in Annäherung zu verwirklichen. Patienten und Therapeuten können, die ethisch gebotene Aufklärung vorausgesetzt, nicht gegen die Gruppenzuordnung verblindet werden.

Immerhin wäre es aber möglich, rehabilitative Leistungen mit nicht-rehabilitativen, also z.B. mit der bei dieser Patientengruppe üblichen ambulanten Behandlung am Wohnort, oder einem Therapieverzicht zu vergleichen. Infrage käme auch ein sog. Wartegruppendesign (Jäckel et al. 1990): ein zufällig ausgewählter Teil der Stichprobe wird sofort behandelt, ein anderer wartet mehr oder weniger unbehandelt einige Wochen bis wenige Monate.

Es ist kritisch hervorzuheben, dass nur eine einzige der von uns identifizierten Studien Rehabilitation mit Nicht-Rehabilitation verglichen hat.

Damit ist sonst keine Studie geeignet, die („absolute“) Wirksamkeit von Rehabilitation zu prüfen. Einem euphemistischen Sprachgebrauch folgend untersuchten alle anderen Studien nicht mehr als „relative Wirksamkeit“, d.h. differentielle Effekte unterschiedlicher Rehabilitationsformen.

Die Frage nach der Wirksamkeit der medizinischen Rehabilitation im Indikationsgebiet „chronische Rückenschmerzen“ im Vergleich zu einer Nicht-Intervention, einem Urlaub, einer

Selbstbehandlung bzw. einer ambulanten Behandlung durch Praxisärzte bleibt somit weiter offen.

3.2.5 Kostenaspekte

Fünf Studien beschäftigen sich – hauptsächlich (IDN 14) oder am Rande (IDN 11,16,19,25) – mit Kostenaspekten der miteinander verglichenen Rehabilitationsformen.

Es dominiert die Kostenträgerperspektive. Zwei Studien weisen auf eine überlegene Effizienz ambulanter Rehabilitationsformen hin; eine zeigt Gleiches für eine stärker beruflich orientierte Rehabilitation (MBOR). Eine methodische orientierte Studie (IDN 16) diskutiert beispielhaft die Schwierigkeiten und Unsicherheiten solcher Studien und damit auch der anderen vier Arbeiten.

3.2.6 Stichprobengrößen und Stichprobenverluste

Berücksichtigen wir alle Arme einer Studie und lassen wir die Verluste, die zwischen der Feststellung der grundsätzlichen Studieneignung (Eligibilität) und der Teilnahme an der letzten Studierhebung auftreten, außer Acht, dann analysierten die Studien zwischen 77 und 850 Personen mit kompletten, auswertbaren Datensätzen. Eine Übersicht über die Stichprobengrößen der eingeschlossenen Studien bietet Tabelle A3 im Anhang.

Die niedrigsten Verlustraten (drop out: 4 % bzw. 6 %) zwischen Studieneintritt und Studienende weisen die beiden Studien auf, die keine Nachbeobachtungszeit vorsahen (IDN 06, 08), also mit der Entlassung abschlossen. Die zwölf Studien mit Nachbeobachtungszeiten zwischen zwei und zehn Monaten haben im Mittel eine Verlustrate von 27 % (SD=12). Die 14 Studien mit einem Follow-Up von zwölf und mehr Monaten zeigen am Ende Verlustraten von im Mittel 36 % (SD=15).

3.2.7 Zusammensetzung der Stichproben

Die Zusammensetzung der Stichproben ist heterogen; es wurden unterschiedliche Ein- und Ausschlusskriterien wirksam. Alle enthalten jedoch einen wenigstens überwiegenden Anteil von Versicherten mit chronischen unspezifischen Rückenschmerzen. Die Sorgfalt der Definitionen „chronisch“ und „unspezifisch“ variiert in einem weiten Bereich.

Gleiches gilt für den Anteil der weiblichen Studienteilnehmer; dieser reflektiert vor allem den jeweiligen Rehabilitationsträger: ehemals LVA-Versicherte mit Rückenschmerzen sind häufiger männlich, ehemals BfA-Versicherte häufiger weiblich. Die im Hintergrund stehende Sozialschichtproblematik wird nur in einer einzigen Studie (IDN 10) ausdrücklich thematisiert.

Nur in zwei Studien sind erkennbar Rehabilitanden unter Kostenträgerschaft der GKV (mit eingeschlossen (IDN 03,11).

3.2.8 Messzeitpunkte und Dauer der Nachbeobachtung

Variabel ist auch die Wahl der Messzeitpunkte; sie variiert zwischen zwei und fünf. Immer erhoben werden Daten vor bzw. zu Beginn und am Ende der Rehabilitation (hierauf begrenzen sich die Studien 06 und 08). 14mal und damit am häufigsten wurden drei Messzeitpunkte realisiert.

Die Nachbeobachtungsdauer liegt zwischen null (zweimal, s.o.) und 24 Monaten (IDN 09). Insgesamt sind es 14 Studien, die das für adäquat gehaltene Follow-Up von mindestens einem Jahr erreichen bzw. einmal überschreiten: die medizinische Rehabilitation sollte nachhaltig nutzen und nicht schaden. Dies kann sicher nicht geprüft werden, wenn die Nachbeobachtung schon nach drei oder weniger Monaten abbricht (was siebenmal der Fall ist).

3.2.9 Nutzenparameter – Outcomes

Auch hier bietet sich ein buntes Bild. Für die Rehabilitation der Sozialversicherungsträger DRV und GKV sind besonders Nutzenparameter zur Abbildung verschiedener Aktivitäts- und Teilhabeaspekte von Bedeutung.

Es gibt nur eine einzige Studie, die nicht auch Schmerz Aspekte berücksichtigte (IDN 21). Rund 80 % der Studien erfassen mit einem breiten Spektrum von Selbstausfüll-Instrumenten Aspekte der Funktionskapazität und der Lebensqualität der Rehabilitanden. Neun verzichten allerdings auf die Erhebung von im engeren Sinne sozialmedizinischen Nutzenmaßen wie z.B. Rückkehr zur Arbeit, Arbeitsunfähigkeit, subjektive Erwerbsprognose, Rentenantragstellung.

Seit längerem wird über mögliche unerwünschte Wirkungen medizinischer Rehaleistungen diskutiert (Höder et al. 2007). Wir fanden sieben Studien (25 %), die dieses Thema zumindest ansprachen.

3.3 Details zur internen Validität der Studien

Tabelle 4 zeigt für jede der berücksichtigten Studien das Ergebnis der Bewertung der internen Validität (= Bewertung des Verzerrungspotenzials), die mit Hilfe der zwölf von der Cochrane Back Review Group (Furlan et al. 2009) vorgeschlagenen Kriterien erfolgte. Obwohl zu den meisten Studien mehrere Publikationen vorliegen (vgl. Tabelle A1 im Anhang), wird verkürzt nur der Erstautor der zeitlich ersten Studienpublikation genannt).

Tabelle 4: Bewertung der internen Validität / Biasrisiko mittels 12 Kriterien (+ = Kriterium erfüllt; ● = Kriterium nicht erfüllt; ? = unklar; Biasrisiko sinkt mit Anzahl erfüllter Kriterien)

IDN	Autorkürzel	Zufällige Zuweisung zu Gruppen	Verdeckte Zuweisung	Verblindung der Studienteilnehmer	Verblindung der Kliniker	Outcome-Beurteilung verblindet	Drop-out beschrieben + akzeptabel	Intention to treat Analyse erfolgt	Keine selektive Berichterstattung	Gruppen zu Beginn vergleichbar	Vergleichbare Kointerventionen	Prozessqualität gesichert	Vergleichbare Messzeitpunkte	Gesamturteil Biasrisiko
01	Bahrke 2006	●	●	●	●	●	?	●	●	+	?	?	+	hoch
02	Bühren 2002	●	●	●	●	●	●	●	+	●	●	?	+	hoch
03	Bürger 2002	●	●	●	●	●	●	●	●	●	?	?	+	hoch
04	Deck 2012	●	●	●	●	●	●	●	+	●	?	?	+	hoch
05	Dibbelt 2006	●	●	?	●	●	●	●	●	+	+	?	+	hoch
06	Franke 2000	+	+	●	●	●	+	●	?	?	+	?	●	hoch
07	Glattacker 2013	●	●	?	●	?	●	●	+	●	?	?	+	hoch
08	Glattacker 2010	●	●	?	●	●	+	●	+	●	?	?	+	hoch
09	Hampel 2009	●	+	?	●	?	●	●	+	●	+	?	+	hoch
10	Hofreuter 2008	●	?	●	●	●	+	●	●	+	+	?	+	hoch
11	Kainz 2006	●	●	●	●	●	●	+	+	●	●	?	+	hoch
12	Kladny 2003	+	?	?	●	?	+	●	+	+	?	?	+	hoch
13	Kleist 2001	+	+	●	●	●	+	●	+	+	●	?	+	niedrig
14	Klingelhöfer 2003	+	?	●	●	●	●	●	+	?	?	?	+	hoch
15	Kolip 2001	●	●	●	●	●	+	●	?	●	?	?	+	hoch
16	Schweikert 2006	+	+	●	●	?	●	?	+	+	?	?	+	hoch
17	Ludwig 2006	●	●	●	●	●	+	●	+	?	+	?	+	hoch
18	Meng 2009	+	+	?	●	?	+	●	+	+	+	?	+	niedrig
19	Mau 2002	●	●	●	●	●	●	●	+	●	?	?	+	hoch
20	Molsberger 2002	+	+	+	●	+	●	+	+	+	+	?	+	niedrig
21	Pomp 2012	●	?	?	?	●	●	●	+	●	+	?	+	hoch
22	Schreiber 2004	+	+	+	?	?	+	●	+	+	?	+	+	niedrig
23	Schwarz 2007	+	+	●	●	?	+	+	+	+	+	?	+	niedrig
24	Stier 2001	●	●	●	●	?	+	●	+	?	+	?	+	hoch
25	Streibelt 2006	+	+	?	●	?	●	●	●	●	?	?	+	hoch
26	Weiß 2013	+	+	●	●	●	+	●	+	+	+	+	+	niedrig
27	Fröhlich 2007	●	●	●	●	●	●	●	+	●	+	?	+	hoch
28	Hüppe 2006	+	+	●	●	●	●	+	+	●	●	?	+	hoch

Sechs von 28 Studien (21 %) weisen nach dieser Bewertungsmethode ein geringes Verzerrungsrisiko bzw. eine hohe interne Validität auf. Die methodisch „beste“ Studie (IDN 20) erfüllte neun der zwölf angewandten Kriterien.

Die meisten Schwierigkeiten bereiteten den Studien die Erfüllung der drei Kriterien zur Verblindung: Keine Studie realisierte eine Verblindung der Behandler, nur eine die Verblindung der Studienteilnehmenden und zwei stellten eine verblindete Erhebung der primären Outcomes sicher (IDN 20,22).

Mit einer Ausnahme erfüllten alle Studien das Kriterium der gleichen Messzeitpunkte für die Outcome-Erhebung und 75 % der Studien ließen keine Anhaltspunkte für selektives Berichten erkennen (z.B. Nichtberichten nichtsignifikanter Ergebnisse).

Die übrigen sieben Kriterien wurden von 14 % bis 43 % der Studien erfüllt.

Fallzahlabeschätzung und Biasrisiko

Nur 10 der 28 Studien berichten von einer systematischen Fallzahlabeschätzung in den berücksichtigten Publikationen.

Fragt man, wie diese 10 Studien im Vergleich zu den übrigen 18 ohne Aussagen zur Fallzahlberechnung mit Blick auf das Biasrisiko abschneiden, zeigt sich Folgendes: Unter den 10 Studien mit angesprochener Fallzahlbegründung erreichen 50 % die Kategorie „geringes Biasrisiko = 6 und mehr erfüllte Kriterien). Unter den übrigen 18 Studien schafft nur eine Studie ein solches Ergebnis ($p=0.006$). Mit einem Mittelwert von 5,8 für die interne Validität zeigen diese Studien einen signifikant besseren Wert als die Studien ohne Nennung der Fallzahl ($M=3,2$; $p=0.001$). Dieses Gütekriterium (zumindest der Berichtsqualität) zeigt sich also mit der internen Validität der Studie assoziiert.

3.4 Zeittrends in der Untersuchungsmethodik

Über die drei definierten Zeiträume (Studienbeginn vor 2000, 2000-2005, nach 2005) ließ sich keine Veränderung in der Zusammensetzung der Studiendesigns erkennen. Insbesondere kam es zu keiner Zunahme des RCT-Anteils (\emptyset 43 %).

Auch das Biasrisiko der Studien veränderte sich über die drei Zeiträume nicht systematisch: die mittlere Anzahl erfüllter Kriterien lag unverändert bei vier.

Wir beobachten keine Zeittrends bei der Anzahl der Studienarme (\emptyset 2), der Größe der Netto-Studienstichproben (\emptyset 300), der Länge der Follow-Up-Zeiträume (Range: Rehabilitationsende bis 24 m; Median 12 m), der Größe der Kohortenverluste (drop out zum letzten Messzeitpunkt \emptyset 30 %, Range: 4-59).

Einen positiven signifikanten zeitlichen Trend finden wir für die Erwähnung eines Ethikvotums: In sechs der acht Studien nach 2005 (75 %) wird der Beratung einer Ethikkommission genannt. In den beiden anderen Zeitperioden liegt der Anteil mit 22 % (2000-2005) und 9 % (vor 2000) deutlich niedriger.

Eine positive Entwicklung zeichnet sich auch für die Diskussion methodischer Limitationen ab (von 55 % über 78 % auf 100%); unverändert (ohne Zeittrend) niedrig liegt die Mitteilung von Ausführungen zu einer systematischen Fallzahlabeschätzung (Ø 36 %).

Kapitel 4 Beantwortung der vier Forschungsfragen

4.1 Wie ist der Stand der Evidenzbasierung von medizinischen Reha-Maßnahmen in Deutschland (beispielhaft am Indikationsfeld Rückenschmerz)?

Obwohl chronische Rückenschmerzen nach wie vor die häufigste Einzelindikation für die medizinische Rehabilitation darstellen und sich weit mehr als 500 Rehabilitationskliniken für dieses Krankheitsbild engagieren, muss die Evidenzbasis weiter in mehrfacher Hinsicht als schwach bezeichnet werden:

- a) Nach wie vor fehlt es an Studien zur „absoluten Wirksamkeit“. Zu den zwei uns bekannten (Jäckel et al. 1990; Hüppe et al. 2006) scheinen jedenfalls in den letzten 14 Jahren keine weiteren hinzugekommen zu sein.
- b) Das bevorzugte „quasi-experimentelle“ Studiendesign ist das der „sequentiellen Kontrollgruppenstudie“. Dieses und verwandte Designs werden gewählt, so ein verbreitetes Argument, weil echte Wirksamkeitsstudien juristisch, ethisch und praktisch nicht machbar wären. Dieses Argument lässt sich juristisch, ethisch und praktisch widerlegen (s.u. Abschnitt 4.4).

Das Hauptproblem des genannten Designs ist, dass es ein Risiko für den sog. Empfänglichkeits-Bias („susceptibility bias“) beinhaltet. Es ist nicht ausgeschlossen, dass die nacheinander gebildeten und am Ende direkt miteinander verglichenen Patientenkohorten sich in prognostischen und für das therapeutische Ansprechen relevanten Merkmalen voneinander unterscheiden (z.B. durch natürliche und jahreszeitliche Schwankungen, veränderte Zuweisungsroutinen der Träger, epidemiologische Trends, gesetzliche Änderungen). Der direkte Vergleich setzt aber voraus, dass es zwischen den Gruppen mit Studienbeginn keine anderen Unterschiede gibt als die Interventionsbedingung. Dies war höchstwahrscheinlich nur bei drei der zehn Studien dieses Zuschnitts der Fall (bei fünf sicher nicht, bei zwei nicht beurteilbar). Um die Verzerrungsmöglichkeit zu korri-

gieren, wenden die Untersucher eine oder mehrere der folgenden Verfahren an: Bildung strukturgleicher Gruppen (Matching), Auswertung in gleich ausgewählten Subgruppen (Stratifizierung), direkte Standardisierung, statistische Adjustierung in multivariaten Modellen. Es liegt auf der Hand, dass parallel gebildete Kontroll- und Kontrastgruppen ein ähnlich hohes (im zweiten Fall höheres) Biasrisiko tragen.

Ein weiteres Verzerrungsproblem („performance bias“) liegt in der wechselseitigen Kontamination der Interventions- und Kontrollbedingung, vor allem, wenn nur eine Klinik mit mehreren Wechseln involviert ist. Es ist nicht sehr plausibel, dass nicht schon die Vorbereitung einer solchen Studie ohne Folgen für das Personal und Klima der Klinik bliebe, und es ist noch unplausibler, dass alles mehrfach zufallsgesteuert in den Ausgangsstatus zurückfallen können soll (Studie 12).

- c) Zu dem eingangs geäußerten Urteil tragen auch unsere Befunde zum akkumulierten Verzerrungsrisiko bei. Nur sechs der 28 Studien erreichen oder überschreiten die von der Cochrane Back Review Group definierte Schwelle von sechs der zwölf Punkte. Der höchste beobachtete Punktwert ist neun und wird von nur einer Studie erreicht (IDN 20), drei Studien erreichen acht Punkte (IDN 22,23,26). Aber auch diese vier relativ belastbarsten Studien sind nicht ohne Probleme: zwei prüften den Zusatznutzen von Akupunktur in einem dreimonatigen Follow-Up (IDN 20,26), eine evaluierte eine neuartigen Form der Rehabilitandenzuweisung (IDN 22). Am überzeugendsten ist die Studie 23 (Vergleich Standardrehabilitation vs. verhaltensmedizinisch orientierte Rehabilitation mit oder ohne telefonische Nachsorge); sie machte sich den Umstand zunutze, dass in einer Klinik zwei organisatorisch und programmatisch getrennte Abteilungen bestanden, denen Patienten einer Quelle zufällig zugewiesen werden konnten. Die Studie muss Einschränkungen in der Qualitätsbeurteilung dadurch hinnehmen, dass weder zur Verblindung der Patienten, noch der Behandler, noch der Beobachter gegenüber der Zuordnung zu einem der Studienarme Stellung genommen wurde. Die Ergebnisse dieser inhaltlich und methodisch besonders relevanten Studie zeigen, dass alle drei Behandlungsbedingungen mit positiven Veränderungen diverser Effektparameter einhergingen, dass die beiden experimentellen Variationen zwar einige statistisch signifikante Zusatznutzen beinhalteten; diese waren aber selten klinisch relevant.
- d) Schließlich widmet sich keine Studie (auch nicht IDN 23) systematisch der Frage, ob die geprüfte und die Standard-Intervention die gewünschte oder vorgeschriebene Prozessqualität realisierte. Dies ist gerade bei hochkomplexen Interventionen wie einer medizinischen Rehabilitation ein besonders kritischer Punkt, der einer eigenen und aufwändigen Prüfung bedarf.

Mit Blick auf die Studienlage, wie sie sich z.B. in zwei systematischen Übersichtsarbeiten zu weiteren Hauptindikationen der von DRV und GKV finanzierten Rehabilitation darstellt

(kardiologische Rehabilitation: Mittag et al 2011; psychosomatische Rehabilitation: Wiegand-Grefe et al 2007), nehmen wir an, dass der von uns berichtete Befund ein einigermaßen repräsentatives Bild auch der Forschungslage zur Evaluation weiterer Teile unseres Systems der medizinischen Rehabilitation gibt.

4.2 Entspricht der methodische Standard der vorliegenden Studien den Ansprüchen, die man für ein HTA anlegen würde?

Diese Frage beziehen wir vor allem auf den Teil eines komplexen HTAs, der sich mit der Wirksamkeit und dem (Netto)Nutzen der interessierenden Technologie beschäftigt. Wir gehen davon aus, dass sich im Kontext eines HTA keine anderen Anforderungen stellen als bei einer Systematischen Übersicht nach Cochrane-, GRADE- oder IQWiG-Standard.

Insofern beantworten wir die Frage mit einem grundsätzlichen Nein. Es fehlt fast vollständig an echten Wirksamkeitsstudien, und die sonst zur Verfügung stehende Evidenz erreicht mehrheitlich nur die Stufe 3 der aktuellen Evidenz-Hierarchie des Oxford Centre of Evidence-based Medicine aus dem Jahr 2011 (OCEBM 2011) oder ist (wie die Akupunkturstudien) bei stärkerem Design inhaltlich peripher im Vergleich zu den zentralen Wirksamkeits- und Nutzenanfragen an unser System der medizinischen Rehabilitation.

Orientieren wir uns an den GRADE Guidelines (zuerst (Guyatt et al. 2011), dann wird man insgesamt nur ein geringes Vertrauen („unser Vertrauen in die Effektschätzung ist begrenzt/ our confidence in the effect estimate is limited“ (Balshem et al. 2011)) in die inhaltlich und statistisch sehr unterschiedlichen Ergebnisse haben können.

Dazu trägt auch die Auswertung bei, die einzelne Verzerrungsrisiken untersuchte und in einem Summenwert zusammenführte (vgl. Tabelle 4), und weiter die breite Vernachlässigung möglicher Schadenrisiken der gewählten Interventionen. Hält man solche für möglich, dann ist es besonders bedauerlich, dass keine weiteren echten Wirksamkeitsstudien vorliegen. In der von uns durchgeführten PETRA-Studie (IDN 28) zeigten sich Hinweise auf unerwünschte Effekte einer medizinischen Rehabilitation bei chronischen Rückenschmerzen im Vergleich zur Fortsetzung der laufenden ambulanten Behandlung.

4.3 Gibt es besondere Lücken oder Schwächen in der Studienlage aus dem deutschen Gesundheitssystem?

Es fehlt an Untersuchungen zur „absoluten“ Wirksamkeit der medizinisch-beruflichen Rehabilitation im Indikationsfeld chronische (unspezifische) Rückenschmerzen. Dies ist manchen Wissenschaftlern durchaus bewusst: So findet sich in der Publikation zu Studie 11 (Vergleich

dreier ambulanter Rehabilitationsformen) die Bemerkung: „Für sich genommen kann die vorliegende Studie keinen Beleg für die Wirksamkeit der drei Angebotstypen erbringen. Prinzipiell kann die Studie die Annahme nicht widerlegen, dass alle drei Therapieangebote gleich unwirksam sind. Dazu fehlt eine Kontrollgruppe mit unbehandelten Patienten.“ Und bei Mangels et al (Studie 23) heißt es: „Eine unbehandelte Kontrollgruppe wäre notwendig gewesen, um die absolute Wirksamkeit der hier bedachten Behandlungszugänge zu evaluieren“ [3].

Auch die Qualität der randomisierten kontrollierten Studien zeigt Verbesserungspotentiale: Von zwölf solcher Studien erreicht nur die Hälfte mindestens sechs der zwölf Qualitätspunkte (einmal neun, dreimal acht, einmal sieben und einmal sechs Punkte) als Hinweis auf ein vergleichsweise geringes Verzerrungspotential. Im Mittel über alle 28 Studien werden 4,1 (SD = 2,2; Median = 3,5) Punkte erreicht.

Die Verzerrungsrisiken der nicht-randomisierten Studien (sequentielles Kontrollgruppendesign, paralleles Kontroll- und Kontrastgruppendesign) werden unterschätzt. Dazu mag auch der gerne benutzte Begriff der „quasi-experimentellen“ Studie beitragen, der eine Designqualität in enger Nachbarschaft zum „experimentellen“ RCT suggeriert.

Die von uns untersuchten Studien sind im Mittel von begrenzter methodischer Qualität; bekannte und vermeidbare Verzerrungsmöglichkeiten werden unzureichend vermieden – erreichbare Anforderungen unzureichend erfüllt. Es fehlt ganz offensichtlich an verbindlichen, jedenfalls an befolgten Standards für Evaluationsstudien im Bereich der Rehabilitation.

Der Qualität der Realisierung der Interventions- und der Kontrollbedingung wird so gut wie keine Aufmerksamkeit geschenkt. Auch werden nicht immer die Zeit- und Aufmerksamkeitsdifferenzen ausgeglichen, die sich aus der Einführung neuer Rehabilitationsleistungen oder -programme ergeben.

Nur eine einzige Studie verfolgte die eingeschlossenen Versicherten über 24 Monate [4]. Die Hälfte der Studien blieb unter einem Follow-up von zwölf Monaten.

Eine systematische Weiterentwicklung von Studienfragen und Studienmethodologie ist nicht erkennbar, auch nicht im Gestaltungsbereich der als stark einzuschätzenden Forschergruppen. Eine gewisse Ausnahme bildet die Gruppe um Prof. Greitemann in Bad Rothenfelde und Münster. Sie hat kumulativ mehrere Studien zu einem Thema vorgelegt und an der Optimierung der von ihr entwickelten Intervention gearbeitet.

Diese Gruppe ist es auch, die ihre Fragestellungen aus einer Rehabilitationsklinik heraus entwickelt hat. Sonst sind es überwiegend Gruppen mit einer akademischen Basis, die sich

³ „An untreated control group would have been necessary to evaluate the absolute effectiveness of the treatment approaches considered“

⁴ Auch von der Studie 04 liegt inzwischen eine 24-Monats-Katamnese vor (Schramm et al 2014)

rehabilitationsmedizinische Partner gesucht haben bzw. von ihnen oder bestimmten Rehabilitationsträgern beauftragt wurden. Eine wissenschaftliche Aktivierung und Selbstständigkeit einer größeren Zahl von Rehabilitationskliniker bzw. -kliniken hat sich durch die annähernd zehnjährige Förderung der Rehaforschungsverbände offensichtlich nicht erreichen lassen.

Die Pflicht, den Studienplan einer öffentlich-rechtlich verfassten Forschungsethikkommission zur Beratung vorzulegen, scheint lange übersehen worden zu sein, jedenfalls finden sich dazu bei 19 von 28 Studien keine Angaben.

Hier, wie in anderen Fällen, kann nicht sicher zwischen einer eingeschränkten Berichtsqualität und dem Fehlen des entsprechenden Merkmals entschieden werden. So wäre auch die Berichtsqualität in Anlehnung an internationale Standards (CONSORT, STROBE) zu verbessern.

4.4 Welche Empfehlungen können aus der ermittelten Studienlage hinsichtlich eines optimierten ^[5] methodischen Designs zukünftiger Evaluationen abgeleitet werden?

Randomisierte kontrollierte Studien sind in der Forschung zur absoluten Wirksamkeit und zum Nutzen einer medizinischen Rehabilitation weiter sehr selten.

Sie sind auch dort zu selten, wo es um die relative Wirksamkeit und den Zusatznutzen innovativer oder durch bestimmte Leistungen/Module angereicherter Rehabilitationsprogramme geht. Aufgrund tatsächlicher aber auch gewählter Schwierigkeiten wird hier öfter auf schwächere Studiendesigns ausgewichen – in zehn Fällen auf das sequentielle Kontrollgruppendesign, das ausschließlich zur Untersuchung der sog. relativen Wirksamkeit verwandt wurde.

4. 4.1 Empfehlungen zu Studien zur relativen Wirksamkeit

In den Bereich der gewählten Schwierigkeiten gehört die Vermutung, RCTs seien in der Rehabilitation ethisch und juristisch unzulässig und praktisch nicht machbar. Dagegen spricht allein schon die Tatsache, dass doch eine Reihe von RCTs in guter Qualität, erfolgreich und nach bestandener juristischer und ethischer Prüfung durchgeführt wurden.

Zu den tatsächlichen Schwierigkeiten gehört, dass es nahezu unmöglich sein dürfte, zwei deutlich unterschiedliche Reha-Programme in einer Klinik zeitlich parallel durchzuführen, es sei denn, man kann die Gruppen strikt, d.h. auch räumlich (im Speisesaal, der Turnhalle, im

⁵ Hier möchten wir im Vergleich zum Gutachtenauftrag etwas vorsichtiger formulieren. „Ideal“ ließe keine weiteren Verbesserungen zu.

Bewegungsbad etc.) voneinander isolieren. Sonst sind Kontaminationseffekte, Patientenbeschwerden und selbst Gruppenwechsel von Patienten in das für sie attraktivere Programm nicht auszuschließen.

Vor diesem Hintergrund wären Studien hilfreich, die Versicherte einer Quelle parallel auf mehrere separierbare und zufällig ausgewählte Abteilungen oder auf Kliniken mit klar unterschiedlichem Therapieprofil randomisierten. Wir fanden neben der Studie 23 eine weitere (Kluth 2006), die diese Grundidee verwirklichte – allerdings in einem sequentiellen Kontrollgruppendedesign innerhalb einer Klinik mit zwei Abteilungen. Diese Studie musste ausgeschlossen werden, da sich zum Text der psychologischen Dissertation keine dazugehörige Veröffentlichung fand; ein Anschreiben an die ehemalige Doktorandin blieb ohne Antwort. Beide Studien zeigen aber, dass so etwas möglich ist.

Glauht man, beim risikoreichen (s.o.) sequentiellen Kontrollgruppendedesign bleiben zu müssen, dann wäre es notwendig, es durch gezielte Variationen und/oder durch einen Vergleich mit dem stärkeren RCT auf seine Verzerrungspotentiale hin zu untersuchen. Dies ist bisher nicht geschehen.

4.4.2 Empfehlungen zu Studien zur „absoluten Wirksamkeit“

Wichtiger ist aus unserer Sicht aber die Frage nach der absoluten Wirksamkeit einer medizinischen Rehabilitation im gewählten Indikationsbereich. RCTs mit dieser Fragestellung stehen vor einer grundsätzlichen Schwierigkeit:

Sie werden mit dem Augenblick praktisch unmöglich, in dem ein Versicherter einen Antrag auf Rehabilitation gestellt hat. Wir halten es für nahezu ausgeschlossen, dass sich Versicherte mit subjektiver Rehabilitationsbedürftigkeit, subjektivem Rehabilitationswunsch und gestelltem Antrag noch auf die Kontrollbedingung Routineversorgung oder Behandlungsverzicht randomisieren lassen, von einer bewilligten Rehabilitation zurücktreten oder diese für wenigstens sechs, besser zwölf Monate aufschieben.

Entfernt denkbar wäre es, ihnen statt einer Rehabilitation einen gleichlangen privaten Urlaub anzubieten. Allerdings wird man für eine solche Studie wahrscheinlich keine Finanzierung finden.

Eine andere Möglichkeit bietet das schon erwähnte Wartegruppendedesign. Es hat zwei Probleme: einerseits verbietet sich bei objektivem Rehabilitationsbedarf und prinzipiell bewilligter Leistung ein längeres therapieloses Warten; die Studie von Jäckel et al. sah eine Wartezeit von, wie man will: „nur“ oder „immerhin“ acht Wochen vor. Dies schließt eine Prüfung der Nachhaltigkeit möglicher Effekt aus. Andererseits handelt man sich alle Schwierigkeiten ein, die sich wie bei dem sequentiellen Kontrollgruppendedesign aus dem Zeitversatz (s.o.) ergeben können.

So wird man einen RCT zur absoluten Wirksamkeit nur dadurch realisieren können, dass man die Studie beginnt, bevor ein Rehabilitationsantrag gestellt wurde. Dass so etwas möglich ist, haben wir in inzwischen vier separaten Studien gezeigt (Hüppe et al. 2006; Schlademann et al. 2007; Hüppe et al. 2008; Mittag et al. 2014 (zur Veröffentlichung angenommen)). In diese vier Rehabilitationsstudien waren diverse Rehabilitationsträger (GKVen und DRVen) und Kliniken involviert. Mehrfach wurden die Studienpläne ethisch und datenschutz- sowie leistungsrechtlich mit positivem Ausgang geprüft. Damit können wir von einer generellen Machbarkeit ausgehen.

Die Besonderheit dieses von uns vorgeschlagenen Studienansatzes ist, dass er auf Daten der GKV aufsetzt, bevor Versicherte Kontakt zum Rehabilitationsträger DRV suchen.

Dazu werden in einem ersten Schritt in Datenbanken der GKV Personen identifiziert,

- die gleichzeitig Versicherte einer GKV und einer DRV sind,
- an einer definierten und in administrativen Daten identifizierbaren Gesundheitsstörung leiden und
- eine besondere Gefährdung der Erwerbsfähigkeit erkennen lassen. Dies lässt sich z.B. durch eine bestimmte Zahl spezifisch verursachter AU-Tage identifizieren und/oder durch Antworten auf einem postalischen oder internetbasierten Fragebogen prüfen.

In einem zweiten Schritt wären bei den so identifizierten („eligible“) Personen einerseits die Krankheitssituation im Sinne einer Fallsicherung und andererseits die subjektive Rehabilitationsbedürftigkeit und Antragsintention abzuklären. Die Fallsicherung ist bei chronisch-unspezifischen Rückenschmerzen kein Problem; wer Rückenschmerzen angibt, wird sie schon „haben“. Die Fallsicherung kann bei chronisch somatischen und psychischen Erkrankungen ein ärztliches Attest oder ein Zusammenführen weiterer Datenquellen (z.B. Arzneimittelverordnungen) nötig machen.

Da medizinische Rehabilitation eine Antragsleistung ist und nicht „von Amts wegen“ erzwungen werden kann, ist im Vorfeld ebenso abzuklären, ob für eine(n) Versicherte(n) eine Rehabilitation aktuell überhaupt infrage käme. Viele subjektiv wie objektiv rehabilitationsbedürftige Versicherte sind zu einer Antragstellung nicht bereit.

Die Hälfte oder ein anderer Teil der verbleibenden Fälle werden drittens nach erfolgter Aufklärung und Einwilligung auf eine forcierte Beratung zur Rehabilitation randomisiert – im Vertrauen darauf, dass sich ein erheblicher Teil der Ausgewählten und so Beratenen zu einem Rehabilitationsantrag entschließt. Die Kontrollgruppe bleibt ohne spezifische Intervention und damit in der aktuellen ambulanten oder – vermutlich selten – unter keiner Behandlung.

Nach zwölf oder mehr Monaten werden die Veränderungen im Gesundheitszustand bestimmt, teils durch Selbstausfüll-Fragebögen, teils durch die Analyse administrativer Daten

(AU-Fälle und –Zeiten, Rückkehr zur Arbeit, Rentenanträge). Der Vorteil dieser letzten Daten liegt in ihrer nahezu vollständigen Verfügbarkeit.

Soweit die großen Linien dieses Studiendesigns.

4.4.3 Details eines optimierten Designs

Im Folgenden gehen wir nacheinander in knapper Form auf Details eines solchen Studienplans ein und diskutieren zum Abschluss dessen Stärken, Schwächen und Grenzen.

- *Stichprobenbasis, Eligibilität und Stichprobenziehung*

Stichprobenbasis ist die Versichertendatei einer oder mehrerer gesetzlicher Krankenversicherungen. In ihnen wird nach Personen gesucht, die bestimmte Einschlusskriterien erfüllen. Diese beziehen sich auf demographische Merkmale, Versichertenstatus, Erkrankungen unter definierten ICD-Codes und Indikatoren der Teilhabestörung (z.B. eine Mindestanzahl spezifischer AU-Tage in den letzten zwölf Monaten). Diese Gruppe wird in eine eigene Datei eingestellt; die einzelnen Fälle werden pseudonymisiert. Die Gruppe muss umfangreich genug sein, um die statistisch erforderliche Größe der Nettostichprobe erreichen zu können.

- *Vorklärungen*

Da es keinen Sinn macht, Versicherte weiter einzubeziehen, die sich selbst nicht für rehabilitationsbedürftig halten und die in naher Zukunft keine Rehabilitationsleistung beantragen und in Anspruch nehmen würden, ist dieses Ausschlusskriterium z.B. mit Hilfe einer postalischen, telefonischen oder internetbasierten Befragung zu klären – im Rahmen einer ersten kurzen Studieninformation („Beobachtung des weiteren Krankheitsverlaufs unter unterschiedlichen Beratungsbedingungen, mögliche Verbesserung des Behandlungsverlaufs“). Diese ist von der GKV zu verschicken. Mit Hilfe derselben Befragung ist, so weit möglich, eine Fall-sicherung vorzunehmen. Es werden nur solche Personen antworten, die sich für das Projekt interessieren. Ein Vergleich von potentiellen Teilnehmern und früh Ausscheidenden kann auf der Basis von GKV-Daten erfolgen, eventuell nur auf dem Boden aggregierter Daten [6].

- *Komplexe Intervention und Kontrollbedingung*

In diesen Bereich gehören zuerst die Ausgestaltung und Organisation der Beratung, der aktiv zu unterstützenden Antragstellung, der raschen Prüfung der persönlichen und versicherungsrechtlichen Voraussetzungen für eine Rehabilitation, der Bewilligung der Leistung, der Zuweisung zu einer oder mehreren (bestimmten?) Kliniken und einer eventuellen Nachsorge.

⁶ Genau dieses Vorgehen hatten wir in einer fünften erfolgreich abgeschlossenen Studie in Zusammenarbeit mit der Techniker Krankenkasse gewählt. Das Manuskript (Hüppe et al) ist eingereicht und wird zur Zeit begutachtet.

Parallel ist die immer komplexe Rehabilitation in Zusammenarbeit mit der DRV (eventuell auch bestimmten Rehabilitationseinrichtungen) zu definieren und ggf. zu erproben [7]. Die häufig zu beobachtende gleichzeitige Entwicklung, Erprobung und Evaluation in einem Studienansatz ist zu vermeiden. Es sind Qualitätsindikatoren und Meilensteine für die wesentlichen Rehabilitationsprozesse zu bestimmen.

Als Kontrollbedingung gilt der unbeeinflusste weitere Behandlungsverlauf am Wohnort.

- *Patienteninformation und Studieneinwilligung*

Nach Bereinigung der eligiblen Gruppe um Versicherte ohne Rehabilitationsintention, ohne ausreichende Fallsicherung und ohne Studieninteresse sind alle restlichen Personen erneut anzuschreiben, über die Details der Studie zu informieren, um ihre Einwilligung zu bitten und zur Ausgangslage zu befragen.

Die Mitglieder der Kontrollgruppe können in allgemeiner Form über Leistungen zur medizinischen Rehabilitation informiert werden. Ihnen ist eine Rückmeldung über ihren Gesundheitszustand und die Studienergebnisse in Aussicht zu stellen.

- *Randomisierung und Verblindungen*

Die Randomisierung auf die wenigstens zwei Studienarme (ein dritter könnte ein verändertes Rehabilitationsprogramm beinhalten) erfolgt extern durch das wissenschaftliche Institut. Eine Verblindung der Versicherten und ihrer Berater ist selbstverständlich nicht möglich. Eine erste Verblindung könnte für diejenigen organisiert werden, die die eingehenden Rehabilitationsanträge prüfen - auch wenn deren Verzerrungspotential unsicher und in seiner Richtung unbestimmt ist. Es könnten nachfolgend auch die Rehabilitationskliniker gegenüber der Studienzugehörigkeit ihrer Patienten verblindet werden, wenn diese den normalen Zuweisungen sozusagen untergemischt würden. Allerdings beinhaltet dies einen Verzicht auf die übliche Datenerhebung am Ende der Rehabilitation (was u.E. in Kauf genommen werden kann).

- *Beratung zur Rehabilitation, Antragstellung*

Die Mitglieder der Interventionsgruppe werden von der Krankenkasse oder einer gemeinsamen Servicestelle zu einer gesondert zu organisierenden Beratung eingeladen. Diese kann auch telefonisch erfolgen. In ihr geht es darum, den Versicherten den Weg zu einer rasch einzuleitenden Rehabilitation zu ebnen. Ihnen wird, sprechen nicht versicherungsrechtliche und persönliche Gründe dagegen, aktiv zu einer Rehabilitation geraten; das Er-

⁷ Auch bei diesem Design bleibt es bei der bekannten Programmevaluation. Die bisherige dreiwöchige Rehabilitation ist zu komplex und „gestalthaft“, um sie in viele einzeln zu kombinierende und zu prüfende Bestandteile zu zerlegen. Es ist uns auch zweifelhaft, ob eine „Standardreha“, der ein intensives Akupunkturprogramm hinzugefügt wird, noch als Standard plus X angesehen werden darf oder nicht doch schon eine gesamthaft veränderte Rehabilitation darstellt. Diese Erwägung hat eine Rückseite: eine aus vielen je evidenzbasierten Leistungen zusammengesetzte Rehabilitation darf u.E. nicht eo ipso als insgesamt evidenzbasiert angesehen werden.

gebnis der Beratung sollte ein ausgefüllter Rehabilitationsantrag sein. Dies ließ sich in einer Studie mit Rückenschmerzkranken in etwa 70 % der Fälle erreichen (Hüppe et al 2006).

Mit der/n betroffenen DRV/en ist eventuell ein Verfahren zu einer schnellen und im Rahmen des Projekts zustimmenden Prüfung des Antrags zu verabreden. Allerdings würde das die o.g. angedeutete Verblindung unmöglich machen. Eine Entscheidung in die eine oder andere Richtung könnte von der aktuellen Bewilligungsquote im relevanten Indikationsbereich abhängig gemacht werden. Ist sie niedrig, sollte das Risiko zahlreicher Ablehnungen von zur Rehabilitation eingeladenen Versicherter vermieden werden.

- *Wirksamkeit und Nutzenparameter, Erhebungsmethoden und -zeitpunkte*

Die mittel- bis langfristige Wirksamkeit und der Nutzen der Rehabilitation lassen sich vorzugsweise auf der Basis von Fragebogenergebnissen und administrativen Daten beurteilen. Auf den Einschluss sozialmedizinisch relevanter Parameter und Indikatoren einer gelingenden Teilhabe ist Wert zu legen. Hausärztliche Urteile scheinen uns nicht sonderlich reliabel; sie dürften in vergleichbarer Qualität und zu vergleichbaren Zeitpunkten nur schwer zu erreichen sein. Das Follow-Up sollte wenigstens zwölf, besser 18 bis 24 Monate umfassen. Zu empfehlen ist eine frühere Zwischenerhebung nach sechs Monaten, auch zur Pflege der Kohorte. Zur Veränderungsmessung sollten direkte und indirekte Methoden genutzt werden.

- *Schadenrisiken*

Von Anfang an sind Schadenrisiken in Betracht zu ziehen. Nutzt man Lebensqualitätsinventare, lassen sich Nutzenchancen und Schadenrisiken zugleich abbilden. Ähnliches gilt für das Mehr oder Weniger von AU-Fällen und –Tagen. Andererseits sind spezifische Risiken einer Rehabilitation denkbar: Heimweh, finanzielle Mehrausgaben, Irritationen am Arbeitsplatz und in der Familie, Verstärkung dysfunktionaler Kognitionen und Verhaltensweisen.

- *Störeinflüsse und ihre Kontrolle*

Solche sind vor allem für die Kontrollgruppe während der ganzen Beobachtungsdauer in Erwägung zu ziehen. Es kann als sicher angenommen werden, dass ein kleiner Teil von ihr (nach unseren Erfahrungen unter 10 %, einmal bis 20 %) auch eine Rehabilitation beantragt und bewilligt bekommt und/oder – was vermutlich häufiger der Fall ist - dass sich einzelne Versicherte rehabilitative Leistungen in ihrem ambulanten Setting auf eigene Kosten oder auf Kosten der GKV verschaffen.

In der Interventionsgruppe sind das Nicht-Antreten einer bewilligten Rehabilitation (bis 10 %) und die vorzeitige Beendigung zu registrieren.

- *Qualitätssicherung der Rehabilitation*

Es ist darauf zu achten, dass die zu prüfende Rehabilitation einerseits qualitative Mindeststandards zu definierten Zeitpunkten („Meilensteine“) einhält, andererseits nicht bewusst

mehr oder anderes tut als üblich. Prüfgrößen zum Rehabilitationsprozess stehen in der KTL-Dokumentation der Kliniken und aus Fragebogenantworten der Patienten zur Verfügung.

- *Kooperationsvereinbarungen*

Oben sind vier prinzipielle Partner genannt worden. Wenigstens im Dreieck GKV, DRV und wissenschaftliche Einrichtung sind solche Vereinbarungen zu schließen. Sie regeln die wechselseitigen Rechte und Pflichten und formulieren ein gemeinsames Interesse an einer unbeschränkten Publikation der Ergebnisse.

- *Ethik und Recht*

Die Studie ist ohne jeden Zweifel ethikkommissionspflichtig. Ein besonderes und frühzeitig anzugehendes Problem liegt in der Organisation eines für alle akzeptablen Datenschutzes. Jeder Sozialversicherungsträger hat hierfür eigenes Personal, eigene Standards und eigene Routinen. Bei der Beteiligung mehrerer Kranken- und Rentenversicherungen empfiehlt sich die Definition von Federführenden.

Grundsätzlich ist darauf zu achten, dass personenidentifizierende Daten nur bei den Sozialversicherungsträgern gehalten werden und dass andererseits (pseudonymisierte!) individuelle Befragungsergebnisse allein den Wissenschaftlern zugänglich sind.

- *Auswertung*

Neben Standardverfahren zur Auswertung eines RCTs kommen hier Mehrebenenmodelle infrage. Alle drei Auswertungsansätze (intention to treat, per protocol, as actual) sind zu nutzen.

4.4.4 Grenzen und Schwächen des vorgeschlagenen Designs

Eine Achillesferse ist die gegenüber der an sich schon hochkomplexen Rehabilitation noch einmal gesteigerte Komplexität der Intervention. Sie verbindet eine Screeningprozedur mit einem Beratungsschritt, einer Antragsbearbeitung und der Durchführung einer mehrwöchigen Rehabilitation. Der Erfolg der Studie hängt von ihrer generalstabmäßigen Planung und dem Engagement aller beteiligten Einrichtungen und Personen ab. Sie ist und bleibt aber störanfällig.

Aus dieser Komplexität sind zwei Einwände abgeleitet worden: es werde nicht die medizinische Rehabilitation an sich evaluiert, sondern erstens ein sehr viel umfangreicheres Programm, das zweitens vermutlich zu Rehabilitanden führe, die sonst in den Einrichtungen nicht auftauchen würden. Formal ist beides richtig.

Jedoch wird in diesem Design nur etwas standardisiert und damit auch transparent, was jeder Rehabilitation zwischen subjektiven Bewertungen, vielfältigen Beratungen und der

Antragstellung immer schon vorausgeht. Es bleibt bei den üblichen Studien außerhalb des Blickfeldes, ist aber dennoch wirksam.

Der zweite Einwand weist eher auf eine Stärke des Studienansatzes und auf eine Schwäche des üblichen Antragsverfahrens hin. Dieses zeichnet sich durch eine bemerkenswerte Passivität der Rehabilitationsträger aus: sie warten, dass sich Antragsteller von sich aus melden, auch wenn Geist und Wortlaut (§§ 4 und 8) von SGB IX den Trägern ein aktiveres Verhalten nahelegen.

Dennoch ist es möglich, dass sich die Stichprobenzusammensetzung gegenüber dem, was üblich ist, ändert. Dies wäre jeweils neu zu prüfen. Dafür stehen Daten aus dem laufenden Qualitätssicherungsprogramm der DRV zur Verfügung. Gegebenenfalls wird man die Gruppenstruktur der im Rahmen des neuen Designs rekrutierten Rehabilitanden an die der üblichen bzw. bisher beobachteten Rehabilitandengruppen anpassen müssen.

Zusammenfassung

Hintergrund: Die Bereitstellung rehabilitativer Leistungen für behinderte oder von Behinderung bedrohte Menschen gehört zu den Grundfunktionen entwickelter Gesundheitssysteme. Neben der Gesetzlichen Unfallversicherung sind bei uns die Gesetzliche Rentenversicherung und die Gesetzliche Krankenversicherung die wichtigsten Träger der medizinischen Rehabilitation. Muskuloskeletale Erkrankungen sind weiter der häufigste Anlass medizinisch-rehabitativer Leistungen der DRV und GKV; unter ihnen spielen chronisch-unspezifische Rückenleiden eine besondere Rolle (17 % aller Leistungen der DRV). Zur Evaluation der immer noch überwiegend stationären Programme für diese Indikationsgruppe wurden vergleichsweise viele Studien durchgeführt. Die Literaturlage scheint nur für die psychosomatische Rehabilitation besser zu sein.

Das folgende Gutachten beurteilt auf der Basis einer systematischen Sichtung aller zwischen 2000 und 2013 veröffentlichten interventionell-vergleichenden Studien deren methodische Qualität. Es beantwortet vier Forschungsfragen:

1. Wie ist der Stand der Evidenzbasierung von medizinischen Reha-Maßnahmen in Deutschland (am Beispiel der Indikation chronischer Rückenschmerz)?
2. Entspricht der methodische Standard der vorliegenden Studien den Ansprüchen, die man für ein HTA (zur medizinischen Rehabilitation) anlegen würde?
3. Gibt es besondere Lücken oder Schwächen in der Studienlage aus dem deutschen Gesundheitssystem?
4. Welche Empfehlungen können aus der ermittelten Studienlage hinsichtlich eines optimierten methodischen Designs zukünftiger Evaluationen abgeleitet werden?

Das Gutachten widmet sich nicht der systematischen Darstellung der Studienergebnisse.

Material und Methoden: Mit Hilfe einer elektronischen Literaturrecherche in den Datenbanken MEDLINE, PsychINFO, PSYINDEX und EMBASE wurde für den Zeitraum Januar 2000 bis Oktober 2013 nach einschlägigen deutsch- oder englischsprachigen Publikationen gesucht. Die Publikationen sollten Aussagen zu Effekten, Wirksamkeit, Nutzen, Schäden und/oder Effizienz von im Rahmen des deutschen Gesundheitssystems durchgeführten Rehabilitationsprogrammen oder -modulen für Patientinnen und Patienten mit andauernden Rückenschmerzen treffen. Berücksichtigt wurden nur mindestens zweiarmige interventionelle Studien mit mindestens einer Vorher- und einer Nachher-Messung. Die Studien sollten entweder ein randomisiertes kontrolliertes Design aufweisen (zufällige Zuteilung der Patienten zu Interventions- oder Kontrollgruppen) oder einem „quasi-experimentellen“ Ansatz folgend ihre Kontrollgruppen zeitlich versetzt oder parallel zu den Interventionsgruppen beobachten

(sequenzielles oder paralleles Kontrollgruppendesign) ohne zufällige Zuweisung der Teilnehmenden zu einer der Gruppen.

Die elektronische Recherche wurde ergänzt um eine Handsuche in Übersichtsarbeiten zum Thema Rückenschmerz und Rehabilitation sowie in den Abstraktbänden der rehabilitationswissenschaftlichen Kolloquien der DRV in den Jahren 2000 bis 2013.

Publikationen, die Daten aus ein und demselben Studienprojekt berichteten, wurden einer Studien-Identifikationsnummer zugeordnet.

Die Bewertung der methodischen Studienqualität im Sinne der internen Validität folgte den aktuellen Vorschlägen der Cochrane Back Review Group. Zum Einsatz kam eine Liste mit zwölf operational definierten Kriterien. Jedes Kriterium wurde mit den drei Antwortkategorien ja (=erfüllt), nein (=nicht erfüllt) oder unklar (=unzureichende Informationen) bewertet. Je weniger Kriterien eine Studie erfüllen kann, umso höher ist die Wahrscheinlichkeit systematisch verzerrter Ergebnisse („risk of bias“). Einer Studie wird eine hohe interne Validität (und damit ein vergleichsweise geringes Verzerrungsrisiko) zugesprochen, wenn mindestens sechs der zwölf Kriterien als erfüllt bewertet werden. Die Bewertung der einzelnen Kriterien der internen Validität wurde für jede Studie transparent tabellarisch dargestellt.

Zu jeder der ausgewählten Studien wurden neben den Aspekten, die zur Bewertung der internen Validität notwendig waren (z.B. Art des Studiendesigns, Messzeitpunkte, Höhe des Verlustes von Patienten im Studienverlauf), weitere Informationen exzerpiert und dokumentiert. Diese dienten der genaueren Charakterisierung der Studie, einer Einschätzung der Berichtsqualität sowie der Generalisierbarkeit der Studienergebnisse.

Literatursuche, Datenextraktion und Qualitätsbewertung wurden jeweils von zwei unabhängigen Personen vorgenommen, bei Diskrepanzen erfolgte eine Konsensentscheidung.

Zur Bewertung möglicher Entwicklungen über den Beobachtungszeitraum dieses Reviews wurden die Studien nach dem Jahr des Rekrutierungsbeginns in drei Gruppen geteilt: vor 2000, 2000-2005, nach 2005.

Die erfassten Daten und Beurteilungen wurden unter Anwendung des Statistikpaketes IBM SPSS Version 22 ausgewertet.

Ergebnisse: Insgesamt wurden 28 Studien identifiziert, die unsere Einschlusskriterien erfüllten und auf die keines der Ausschlusskriterien zutraf. Dies ist eine bemerkenswert hohe Zahl [⁸]. Fast alle konzentrierten sich auf Rehabilitationsleistungen der Deutschen Rentenversicherung. Ihre Ergebnisse werden in 50 Publikationen berichtet.

⁸ In unserem früheren Review hatten wir für den Zeitraum 1980 bis 2003 bei ähnlicher Suchstrategie nur 24 Studien mit Kontrollgruppe gefunden.

Grundsätzlich ist zwischen Studien zur absoluten vs. relativen Wirksamkeit von medizinischer Rehabilitation bei chronischen Rückenschmerzen zu unterscheiden. Nur eine einzige Studie verglich die Effekte einer standardmäßigen dreiwöchigen stationären Rehabilitation mit ihrem Gegenteil, der Nicht-Rehabilitation, hier in Form der weiterlaufenden ambulanten Behandlung.

Die 27 weiteren Studien verglichen unterschiedliche Rehabilitationsformen und -leistungen bzw. -module miteinander. Sie können damit keine Stellung zu der Frage nehmen, ob unser System der rehabilitativen Versorgung für sich genommen „absolut“ wirksam ist. Sie widmeten sich der sog. relativen Wirksamkeit der von ihnen geprüften Interventionen, untersuchten deren differentielle oder Zusatzeffekte.

Zwölf der 28 Studien folgten dem Design der randomisierten-kontrollierten Studie (RCT); 16 können als „quasi-experimentell“ bezeichnet werden; zehn von ihnen erscheinen im Gewand des sequentiellen Kontrollgruppendesign, sechs beobachteten ihre Interventions- und Vergleichsgruppen parallel; unter ihnen rekrutierten vier „Kontrastgruppen“, die sicher aus einer anderen Grundgesamtheit stammten.

Beurteilt man die Qualität aller Studien mit Hilfe eines international eingeführten Werkzeugs auf ihre Verzerrungsrisiken, dann werden im Mittel vier von zwölf möglichen Punkten erreicht. Die beste Studie, ein RCT, kam auf neun Punkte; diese Studie widmete sich wie weitere als überdurchschnittlich valide zu beurteilende Studien dem Vergleich der Standardrehabilitation mit einem, um eine abgrenzbare Leistung angereicherten Programm (in diesem Fall durch Akupunktur).

Von den zwölf RCTs erreichten oder überschritten sechs die Grenze höherer Validität (≥ 6 Punkte); dies gelang keiner der nicht-randomisierten Studien.

Am inhaltlich wie methodisch überzeugendsten ist eine randomisierte kontrollierte Studie (IDN 23) zum Vergleich der stationären Standardrehabilitation mit einem ebenfalls stationären verhaltensmedizinisch orientierten Programm mit oder ohne telefonische Nachsorge; sie erreicht ebenfalls acht Punkte und zeigt, was im Bereich der Rehabilitationsforschung methodisch möglich ist. Verbesserungswürdig war hier vor allem die Behandlung der Verblindungsproblematik.

Erhebliche Validitätsprobleme resultieren aus den durchschnittlich hohen Kohortenverlusten: Bei Studien mit einer zwölf-monatigen Nachbeobachtungszeit liegen sie im Mittel bei 36 %. Dies scheint als unvermeidbar hingenommen zu werden. Gezielte Bemühungen oder – weitergehend – experimentelle Ansätze zu einer intensiveren Kohortenpflege und Studienadhärenz sind nicht zu verzeichnen. Eine Ausnahme bildet wieder die Studie 23 mit nur 6 % Verlust über zwölf Monate; möglicherweise verdankt sich dies einer telefonischen Ankündigung des Abschlussfragebogens.

Problematisch ist aus unserer Sicht auch, dass die Prozessqualität jedenfalls der zu prüfenden Intervention nicht systematisch untersucht und bewertet wird. Wir finden nur zwei Studien, die hierzu gezielt Angaben machen. Und so bleibt auch die Frage nach möglichen Änderungen in den Routineprogrammen meist unbeantwortet: Blieb es in jedem Fall bei der Standardintervention „nach den Vorgaben der Rehabilitationsträger“, wenn diese entweder als Kontrollbedingung oder als Basis einer angereicherten Rehabilitation diente?

Im Bereich der Auswertung fällt auf, dass nur vier Studien ihre Daten nach dem Intention-to-treat-Prinzip ausgewertet haben.

Kostenaspekte wurden von fünf Studien untersucht. Es dominierte die Kostenträgerperspektive. Es zeigte sich eine überlegene Effizienz ambulanter Verfahren sowie einer stärker beruflich orientierten Rehabilitationsform.

Wesentliche Veränderungen im Bereich der Studiendesigns und Studienqualitäten waren im Vergleich der drei gewählten Zeiträume des Studienbeginns nicht zu beobachten.

Diskussion:

Mit Blick auf die Studienlage, wie sie sich in systematischen Übersichtsarbeiten zu weiteren Hauptindikationen der von DRV und GKV finanzierten Rehabilitation darstellt (kardiologische Rehabilitation; psychosomatische Rehabilitation) nehmen wir an, dass unser Gutachten ein einigermaßen repräsentatives Bild auch der Forschungslage zur Evaluation weiterer Teile unseres Systems der medizinischen Rehabilitation gibt.

Bei der Beurteilung der methodischen Qualität aller Studien ist zu berücksichtigen, dass einer der zwölf erreichbaren Punkte für die Randomisierung selbst und vier weitere für mögliche Verblindungsschritte vergeben werden können. So wundert es nicht, dass keine der sog. quasi-experimentellen Studien die 6-Punkte Grenze erreicht. Eine Verblindung der Rehabilitanden gegenüber ihrer Exposition ist bei Studien zur absoluten Wirksamkeit ausgeschlossen und erfordert bei solchen zur relativen Wirksamkeit besondere Vorkehrungen zur Verhüllung der Studieninhalte bzw. -ziele und zur Trennung der Gruppen, wenn nur eine Klinik involviert ist.

Die Verzerrungsrisiken der „quasi-experimentellen“ Studien, hier vor allem der Studien nach dem sequentiellen Kontrollgruppendesign werden unterschätzt, sie bleiben jedenfalls undiskutiert.

Beantwortung der vier Gutachtenfragen, eine dringende Empfehlung:

Zum Stand der Evidenzbasis: Beurteilt man die methodische Qualität der Studien und die durch sie vermittelte Ergebnissicherheit im Bereich der medizinischen Rehabilitation für

chronisch Rückenschmerzkrankte unter Berücksichtigung ihrer Verzerrungsrisiken, Schätzgenauigkeiten und generellen Vertrauenswürdigkeit, dann wird man die gesamte Evidenzbasis, legt man die sechsstufige Schulnotenskala zugrunde, nur als „befriedigend“ einstufen können. Die Studienlage zur absoluten Wirksamkeit ist „ungenügend“.

Zu HTA-Standards: Ein Grundproblem für ein rehabilitationsspezifisches HTA wäre, dass man kaum drei Studien zu einem Thema findet. Überzeugende Hinweise auf eine kumulative Forschung – eine Studie geht aus der anderen hervor, man arbeitet sich zunehmend tiefer in ein umschriebenes Forschungsgebiet hinein – finden sich für unser Thema nur in Ansätzen. Studien zu isolierten Interventionen bzw. Themen überwiegen, auch in Zentren, die nach unserer Kenntnis sonst kumulativ forschen.

Nur wenige Studien scheinen ihren Ausgang von einer HTA-typischen „policy question“ genommen zu haben. Es überwiegen Studien, denen klinische und/oder wissenschaftliche Fragen zugrunde liegen. Es wäre daher überraschend, wenn die Studienergebnisse eine wesentliche praktische Bedeutung gehabt hätten. In kaum einem Fall scheinen sie zu einer Umgestaltung der Verwaltungs- und Organisationsroutinen der Rehabilitationsträger geführt zu haben, wenn man von der zunehmenden Betonung der ambulanten Rehabilitationsformen in unserem Indikationsbereich absieht. Dies wäre z.B. bei den Ergebnissen der beiden Untersuchungen zu Zuweisungsmodalitäten sowie der einen zum Nachsorgeprogramm zu erwarten gewesen. Andererseits ist nicht zu übersehen, dass sich die Rehabilitationsprogramme für chronisch Rückenschmerzkrankte in den letzten beiden Dekaden fortentwickelt haben, durchaus in Richtung Rehabilitandenzentrierung und Verhaltensmedizin.

Studien zur Kosteneffizienz, wie sie in einem typischen HTA zu berücksichtigen wären, sind insgesamt selten; die auffindbaren sind deskriptiv und berücksichtigen nur die Kostenträgersperspektive. Interventionelle Studien zu Erhöhung der Effizienz fehlen.

Lücken und Schwächen der Studienlage: Standards zu evaluativen Rehabilitationsstudien sollten besonders auf eine Vereinheitlichung von Ein- und Ausschlusskriterien, eine angemessene Fallzahlplanung, erweiterte Nachbeobachtungszeiten, Mindestanforderungen an sozialmedizinisch relevante Effektparameter, eine einheitliche Berichtsqualität auch zur Beurteilung der Prozessqualität und auf eine Vorlage der Studienpläne bei einer öffentlich-rechtlich verfassten Forschungsethikkommission hinwirken. Damit könnten bekannte Verzerrungsmöglichkeiten vermieden und erreichbare Anforderungen umfassender erfüllt werden. Dies könnte dazu beitragen, in weiteren Zeitperioden zu einer kumulativen Fortentwicklung im Bereich der bearbeiteten Studienfragen, der Studiendesigns in Richtung RCTs und der internen Validität der Studien voranzukommen. Es wäre hierzu hilfreich, zu einer engeren Verbindung von klinischer Rehabilitationspraxis und klinischer Epidemiologie (als Grundlagenwissenschaft einer evidenzbasierten Rehabilitation) zu finden. Die Kombination von

Rehabilitationspraxis und Forschungsexpertise unter einem Dach ist selten. Es überwiegen Forschungskonsortien mit einer akademischen Basis, die sich rehabilitationsmedizinische Partner gesucht haben bzw. – in wenigen Fällen - von bestimmten Rehabilitationsträgern beauftragt wurden. Eine wissenschaftliche Aktivierung und Selbstständigkeit einer größeren Zahl von Rehabilitationsklinikern bzw. -kliniken hat sich durch die rund zehnjährige Förderung von acht Rehabilitationsforschungsverbänden offensichtlich nicht erreichen lassen.

Eine dringende Empfehlung richtet sich auf die notwendige Bearbeitung der Frage nach der sog. absoluten Wirksamkeit der bei uns üblichen medizinischen Rehabilitation. Hierzu wird ein schrittweise entwickeltes Studiendesign vorgeschlagen, das geeignet scheint, faktische und gewählte Grenzen der bisherigen Forschung zu überwinden – nicht nur im Bereich der rüschenschmerzbezogenen Verfahren.

Literaturverzeichnis

(Die Publikationen zu den im Text analysierten Studien sind extra im Anhang in den Tabellen A1 und A2 aufgeführt.)

- Bai A, Shukla V, Bak G, Wells G. Quality Assesment Tools Project Report. Ottawa: University; 2012.
- Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol* 2011; 64: 401-406.
- Dreier M, Borutta B, Stahmeyer J, Krauth C, Walter U. Vergleich von Bewertungsinstrumenten für die Studienqualität von Primär- und Sekundärstudien zur Verwendung für HTA-Berichte im deutschsprachigen Raum. Köln: University; 2010.
- Furlan A, Pennick V, Bombardier C, van Tulder M. 2009 updated method guidelines for systematic reviews in the Cochrane Back Review Group. *Spine (Phila Pa 1976)* 2009; 34: 1929-1941.
- Guyatt G, Oxman A, Akl E, Kunz R, Vist G, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 2011; 64: 383-394.
- Higgins J, Green S. *Cochrane Handbook for Systematic Reviews of Interventions* 2011. <http://handbook.cochrane.org/>. Zugriffsdatum: 03.03.2014
- Höder J, Mittag O, Präcklein C, Raspe H. Risiken und Nebenwirkungen von medizinischer Rehabilitation: Annäherungen an ein wenig beachtetes Thema. *Zeitschrift für Allgemeinmedizin* 2007; 83: 153-158.
- Hüppe A, Glaser-Möller N, Raspe H. Trägerübergreifendes Projekt zur Früherkennung von Rehabilitationsbedarf bei Versicherten mit muskuloskelettalen Beschwerden durch Auswertung von Arbeitsunfähigkeitsdaten: Ergebnisse einer randomisierten, kontrollierten Evaluationsstudie. *Gesundheitswesen* 2006; 68: 347-356.
- Hüppe A, Parow D, Raspe H. Wirksamkeit und Nutzen eines Screeningverfahrens zur Identifikation von rehabilitationsbedürftigen Personen mit Diabetes mellitus Typ 2: eine randomisierte, kontrollierte Evaluationsstudie unter Versicherten der Hamburg Münchener Krankenkasse. *Gesundheitswesen* 2008; 70: 590-599.
- Hüppe A, Raspe H. Die Wirksamkeit stationärer medizinischer Rehabilitation in Deutschland bei chronischen Rückenschmerzen: eine systematische Literaturübersicht 1980-2001. *Rehabilitation* 2003; 42: 143-154.
- Hüppe A, Raspe H. Zur Wirksamkeit von stationärer medizinischer Rehabilitation in Deutschland bei chronischen Rückenschmerzen: Aktualisierung und methodenkritische Diskussion einer Literaturübersicht. *Rehabilitation* 2005; 44: 24-33.
- Jäckel W, Cziske R, Gerdes N, Jacobi E. Überprüfung der Wirksamkeit stationärer Rehabilitationsmaßnahmen bei Patienten mit chronischen Kreuzschmerzen: eine prospektive randomisierte, kontrollierte Studie. *Rehabilitation* 1990; 29: 129-133.
- Klemperer D. Qualität und Qualitätskontrolle in der Medizin. Patienten im Gesundheitssystem - Patientenunterstützung und -beratung. Damkowsky W GS, Luckey K. Augsburg, Marco-Verl.; 1995: 189-216.
- Kluth W. Wirksamkeitsvergleich zwischen verhaltensmedizinisch orientierter orthopädischer Rehabilitation und einer Standardrehabilitation [Dissertation]. Hamburg: Universität Hamburg; 2006.
- Lange S, Thomas S. Anforderungen an die Evidenz für Systementscheidungen zur Aufnahme medizinischer Maßnahmen in den Leistungskatalog der Gesetzlichen Krankenversicherung. *Z Evid Fortbild Qual Gesundhwes* 2010; 104: 642-647.
- Meyer T, Gutenbrunner C, Bickenbach J, Cieza A, Melvin J, et al. Towards a conceptual description of rehabilitation as a health strategy. *J Rehabil Med* 2011; 43: 765-769.
- Mittag O, Döbler A, Pollmann H, Farin-Glattacker E, Raspe H. Praktikabilität und Nutzen eines aktiven Screenings auf Rehabedarf mit anschließender schriftlicher Beratung zur Rehaantragstellung bei AOK-Versicherten im Disease-Management-Programm Diabetes Typ 2 (PARTID-Studie). *Rehabilitation* 2014 (zur Veröffentlichung angenommen);
- Mittag O, Schramm S, Böhmen S, Hüppe A, Meyer T, Raspe H. Medium-term effects of cardiac rehabilitation in Germany: systematic review and meta-analysis of results from national and international trials. *European Journal of Cardiovascular Prevention & Rehabilitation* 2011; 18: 587-593.
- Moher D, Hopewell S, Schulz K, Montori V, Gotzsche P, et al. CONSORT 2010 Explanation and Elaboration: Updated guidelines for reporting parallel group randomised trials. *J Clin Epidemiol* 2010; 63: e1-37.

- MRC. Developing and evaluating complex interventions: new guidance. 2008.
<http://www.mrc.ac.uk/Utilities/Documentrecord/index.htm?d=MRC004871>. Zugriffsdatum:
26.2.2014
- OCEBM. The Oxford 2011 Levels of Evidence. 2011.
http://www.cebm.net/mod_product/design/files/CEBM-Levels-of-Evidence-2.1.pdf.
Zugriffsdatum: 27.2.2014
- Raspe H. Evidence based medicine: Modischer Unsinn, alter Wein in neuen Schläuchen oder aktuelle
Notwendigkeit? Z Arztl Fortbild 1996; 90: 553-562.
- Raspe H. Rückenschmerzen. Berlin: Robert Koch-Institut; 2012.
- Schlademann S, Hüppe A, Raspe H. Ergebnisse einer randomisierten kontrollierten Studie zur
Akzeptanz und zu Outcomes einer Beratung auf stationäre medizinische Rehabilitation unter
erwerbstätigen GKV-Versicherten mit rheumatoider Arthritis. Gesundheitswesen 2007; 69:
325-335.
- Schramm S, Hüppe A, Jürgensen M, Deck R. Begleitete Eigeninitiative nach der Reha ("Neues
Credo") - Langzeitergebnisse der quasiexperimentellen Interventionsstudie. Rehabilitation,
online first 07.01.2014 DOI: 10.1055/s-0033-1358388
- Shadish W, Cook T, Campbell D. Experimental and Quasi-Experimental Designs for Generalized
Causal Inference Boston: Houghton Mifflin Company 2002.
- SIGN. Critical appraisal: Notes and checklists. 2013.
<http://www.sign.ac.uk/methodology/checklists.html>. Zugriffsdatum: 03.03.2014
- von Elm E, Altman D, Egger M, Pocock S, Gøtzsche P, et al. The Strengthening the Reporting of
Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting
observational studies. Lancet 2007; 370: 1453-1457.
- Wiegand-Greve S, Schmid-Ott G, Meermann R, Jacobi C, Paar GH, Lamprecht F. Ergebnisse einer
Evaluation aller Wirksamkeitsstudien psychosomatischer Rehabilitation. Psychotherapie 2007,
12: 36-45.
- WHO. World Report on Disability. Malta: WHO; 2011.

Tabelle A1: Aufstellung der eingeschlossenen Volltexte mit ihrer Zuordnung zu den Identifikationsnummern (IDN) der Studien

IDN	Kürzel	Literaturangabe
01	Bahrke 2006	Bahrke U, Bandemer-Greulich U, Fikentscher E, Müller K, Schreiber B, Konzag TA: Chronischer Rückenschmerz mit suppressiver Schmerzverarbeitung – Zur Optimierung des Rehabilitationserfolgs einer bislang vernachlässigten Rehabilitandengruppe. Rehabilitation 2006; 45: 336-344
		Bandemer-Greulich U, Bosse B, Fikentscher E, Konzag TA, Bahrke U: Wirksamkeit psychologischer Interventionen auf die Schmerzverarbeitung innerhalb einer orthopädischen Rehabilitation von chronischen Rückenschmerzen. Psychother Psych Med. 2008; 58: 32-37
		Bosse B, Fikentscher E, Bandemer-Greulich U, Müller K, Bahrke U: Multimorbide Rückenschmerzpatienten – Charakterisierung und Erfolg innerhalb der stationären orthopädischen Rehabilitation. Phys Med Rehab Kuror 2007; 17: 197-202
02	Bührlen 2002	Bührlen B, Jäckel WH: Teilstationäre orthopädische Rehabilitation: Therapeutische Leistungen, Behandlungsergebnis und Kosten im Vergleich zur stationären Rehabilitation. Rehabilitation 2002; 41:148-59.
03	Bürger 2002	Bürger W, Dietsche S, Morfeld M, Koch U: Ambulante und stationäre orthopädische Rehabilitation - Ergebnisse einer Studie zum Vergleich der Behandlungsergebnisse und Kosten. Rehabilitation 2002; 41: 92-102
04	Deck 2012	Deck R, Schramm S, Hüppe A: Begleitete Eigeninitiative nach der Reha („neues Credo“) – ein Erfolgsmodell? Rehabilitation 2012; 51: 316-325
05	Dibbelt 2006	Greitemann B, Dibbelt S, Büschel C: Integriertes Orthopädisch-Psychosomatisches Konzept zur medizinischen Rehabilitation von Patienten mit chronischen Schmerzen des Bewegungsapparates – Langfristige Effekte und Nachhaltigkeit eines multimodalen Programmes zur Aktivierung und Umorientierung. Zeitschrift für Orthopädie und ihre Grenzgebiete 2006; 144: 255-66
		Dibbelt S, Greitemann B, Büschel C: Nachhaltigkeit orthopädischer Rehabilitation bei chronischen Rückenschmerzen – Das Integrierte orthopädisch-psycho-somatische Behandlungskonzept (IopKo). Rehabilitation 2006; 45: 324-35
06	Franke 2000	Franke A., Gebauer S, Franke K, Brockow T: Akupunktmassage nach Penzel versus klassische Teilmassage und Einzel- versus Gruppenkranken-gymnastik bei chronischen Rückenschmerzen - eine randomisierte, kontrollierte klinische Studie in 2x2-faktoriellem Design Forsch Komplementärmed Klass Naturheilkd 2000; 7: 286-293
07	Glattacker 2013	Glattacker M, Dudeck A, Dibbelt S, Quatmann M, Greitemann B, Jäckel WH: Evaluation einer Intervention zur partizipativen Vereinbarung von Rehabilitationszielen bei Patienten mit chronischen Rückenschmerzen. Rehabilitation 2013; 52: 257-265
08	Glattacker 2010	Glattacker M, Heyduck K, Meffert C: Illness beliefs, treatment beliefs and information needs as starting points for patient information – evaluation of an intervention for patients with chronic back pain. Patient Education and Counseling 2012; 86: 378-89
		Glattacker M, Heyduck K, Meffert C: Illness beliefs and treatment beliefs as predictors of short-term and medium-term outcome in chronic back pain. J Rehabil Med. 2013; 45: 268-76
		Glattacker M, Heyduck K, Meffert C: Entwicklung, Implementierung und Umsetzung einer Intervention zur bedarfsgerechten Patienteninformation für Rehabilitandinnen und Rehabilitanden. Rehabilitation 2010; 49: 292-300
09	Hampel 2009	Hampel P, Gräf T, Krohn-Grimberghe B, Thomsen M, Mohr B: Effektivität eines kognitiv-behavioralen Depressionsbewältigungstrainings in der stationären orthopädischen Rehabilitation bei chronisch unspezifischem Rückenschmerz und Depressivität. Zeitschrift für Klinische Psychologie und Psychotherapie 2009; 38: 154-65
		Hampel P, Graef T, Krohn-Grimberghe B, Tlach L: Effects of gender and cognitive-behavioral management of depressive symptoms on rehabilitation outcome among inpatient orthopedic patients with chronic low back pain: a 1 year longitudinal study. Eur Spine J 2009;18: 1867-1880
		Mohr B, Krohn-Grimberghe B, Gräf T, Schulze J., Petermann F, Hampel P: Patienten mit chronisch unspezifischem Rückenschmerz: Zur Bedeutung psychosozialer Merkmale. Rehabilitation 2009; 48: 288 – 297
		Tlach L, Hampel P: Long term effects of a cognitive-behavioral training program for the management of depressive symptoms among patients in orthopaedic inpatient rehabilitation of chronic low back pain: a 2 year follow up. Eur Spine J 2011, 20:2143-2151

10	Hofreuter 2008	Hofreuter K, Koch U, Morfeld M: Die Bedeutung sozialer Ungleichheit als Prädiktor für die berufliche Wiedereingliederung von chronischen Rückenschmerzpatienten nach medizinischer Rehabilitation. Gesundheitswesen 2008; 70: 145-53
		Morfeld M, Koch U: Optimierung des Rehabilitationserfolges bandscheibenoperierter Patienten durch ein ergänzendes kognitiv-verhaltenstherapeutisches Behandlungsprogramm. Abschlussbericht zum Projekt B4 im Norddeutschen Verbund für Rehabilitationsforschung, Universitätsklinikum Hamburg-Eppendorf,
11	Kainz 2006	Kainz B, Gülich M, Engel EM, Jäckel H: Vergleich dreier ambulanter Therapieformen zur Behandlung chronischer Rückenschmerzen – Ergebnisse einer multizentrischen, clusterrandomisierten Studie. Die Rehabilitation 2006; 45: 65-77
12	Kladny 2003	Kladny B, Fischer FC, Haase I: Wertigkeit der muskulären segmentalen Stabilisierung zur Behandlung von Rückenschmerz und Bandscheibenerkrankungen im Rahmen der ambulanten Rehabilitation. Z Orthop 2003; 141: 401–405
13	Kleist 2001	Kleist B, Alliger K, Winter S, Beyer WF: Work Hardening bei chronisch unspezifischen Rückenschmerzen in der stationären Rehabilitation. Praxis Klinische Verhaltensmedizin und Rehabilitation 2001; 54: 145-150
		Winter S: Evaluation des Work Hardening bei chronischen unspezifischen Rückenschmerzen. Eine empirische Vergleichsstudie. Dissertation, München 2002
		Kleist B, Winter S, Alliger K. Work Hardening – Evaluation und Implementation des Work-Hardening Programms bei Patienten mit chronisch unspezifischen Rückenschmerzen in der stationären medizinischen Rehabilitation. Abschlussbericht des Projektes B1 im Rehabilitationswissenschaftlichen Forschungsverbund Bayern, Rheumaklinik Bad Füssing.
14	Klingelhöfer 2003	Klingelhöfer HE, Timm A: Wirtschaftlichkeit ambulanter Rehabilitation – Abschließende Ergebnisse eines Projektes in Mecklenburg-Vorpommern. Rehabilitation 2005; 44: 1-13
		Klingelhöfer HE, Lätzsch A: Wirtschaftlichkeitsvergleich ambulante versus stationäre Rehabilitation. Gesundheitswesen 2003; 65: 163-166
		Matschke J, Klingelhöfer HE, Lätzsch A: Wirtschaftlichkeitsparameter beim Auf- und Ausbau von Einrichtungen ambulanter Rehabilitation für die gesetzliche Rentenversicherung. Abschlussbericht des Projektes A3 im Forschungsverbund Sachsen-Anhalt/ Mecklenburg-Vorpommern. Universität Greifswald
15	Kolip 2001	Kolip P, Czujek J, Greitemann G, Rosowski E, Schmidt B, Slangen K: „Lebenslust statt Krankheitsfrust“ – Implementation und Evaluation eines Programms zur aktiven Patientenbeteiligung in der stationären Rehabilitation chronischer RückenschmerzpatientInnen. Rehabilitation 2001; 40: 267-74
		Slangen K, Kolip P, Schmidt B, Rosowski E, Czujek J, Greitemann B: Aktive Patientenbeteiligung in der Rehabilitation. Evaluation eines psychologisch-somatischen Aktivierungstrainings für Rückenschmerzpatienten. Weinheim: Juventa 2002
16	Schweikert 2006	Leidl R, Jacobi E, Knab J, Schweikert B: Entscheidungen bei „schwierigen“ Kosteneffektivitätsrelationen am Beispiel eines klinischen Versuchs aus der Rehabilitation. Gesundheitswesen 2006; 68: 249-256
		Schweikert B, Jacobi E, Seitz R, Cziske R, Ehler A, Knab J, Leidl R: Effectiveness and cost-effectiveness of adding a cognitive behavioral treatment to the rehabilitation of chronic low back pain. Journal Rheumatol. 2006; 33: 2519-26.
		Leidl S, Schweikert B. Projekt 4: Gesundheitsökonomische Analysen. Endbericht. Rehabilitationswissenschaftlicher Forschungsverbund Ulm, Universität Ulm
17	Ludwig 2006	Ludwig FJ, Unlü AI, Melzer C, Daalman HH: Effekte der stationären medizinischen Rehabilitation bei Rückenschmerzpatienten. Zeitschrift für Orthopädie und ihre Grenzgebiete 2006; 144: 569-76
18	Meng 2009	Meng K, Seekatz B, Rosband H, Worrigen U, Vogel H, Faller H: Intermediate and long-term effects of a standardized back school for inpatient orthopedic rehabilitation on illness knowledge and self-management behaviors: a randomized controlled trial. The Clinical Journal of Pain 2011; 27: 248-57
		Meng K, Seekatz B, Rosband H, Worrigen U, Vogel H, Faller H: Entwicklung eines standardisierten Rückenschulungsprogramms für die orthopädische Rehabilitation. Rehabilitation 2009; 48: 335-44
19	Mau 2002	Mau M, Merkesdal S, Busche T, Bauer J: Prognose der sozialmedizinischen Entwicklung ein Jahr nach teilstationärer oder stationärer Rehabilitation wegen Dorsopathie. Rehabilitation 2002; 41: 160-166

		Merkesdal S, Bernitt K, Busche T, Bauer J, Mau W: Gegenüberstellung der Krankheitskosten im Jahr vor und nach stationärer und ambulanter Rehabilitation bei Personen mit Dorsopathien. Rehabilitation 2004; 43: 83-89
20	Molsberger 2002	Molsberger AF, Mau J, Pawelec DB, Winkler J: Does acupuncture improve the orthopedic management of chronic low back pain – a randomized, blinded, controlled trial with 3 months follow up. Pain 2002; 99: 579-87
21	Pomp 2013	Pomp S, Fleig L, Schwarzer R, Lippke S: Effects of a self-regulation intervention on exercise are moderated by depressive symptoms: a quasi-experimental study. International Journal of Clinical and Health Psychology 2013; 13: 1-8
22	Schreiber 2004	Schreiber B, Bandemer-Greulich U, Uhlemann K, Müller K, Müller-Pfeil J, Kreutzfeldt A, Fikentscher E, Bahrke U: Behandlungsspezifisch beim chronischen Rückenschmerz: Ist die optimierte Rehabilitationszuweisung ausreichend? Rehabilitation 2004; 43: 142-51
		Fikentscher E. Optimierung der Rehabilitationszuweisung beim chronischen Rückenschmerz. Abschlussbericht Projekt C1, Forschungsverbund Rehabilitationswissenschaften Sachsen-Anhalt/Mecklenburg-Vorpommern, Universität Halle
23		Schwarz S, Mangels M, Sohr G, Holme M, Worringen U, Rief W: Welche klinische Bedeutung hat die Vorabzuweisung von Patienten zur orthopädischen bzw. verhaltensmedizinisch orthopädischen Rehabilitation durch den Rentenversicherungsträger? Rehabilitation 2007; 46: 323-32
		Schwarz S: Untersuchungen zur Optimierung der Zuweisung von Patienten zur stationären klassisch-orthopädischen oder verhaltensmedizinisch-orthopädischen Rehabilitation. Dissertation, Philipps Universität Marburg; 2008.
		Schwarz S, Mangels M, Sohr G, Holme M, Worringen U, Rief W: Patienten mit vs. ohne psychische Störung in der orthopädischen Rehabilitation Schmerz 2008; 22: 67-74
		Mangels M: Implementierung verhaltensmedizinischer Ansätze in der stationären orthopädischen Rehabilitation. Dissertation, Philipps Universität Marburg; 2008
		Mangels M, Schwarz S, Worringen U, Holme M, Rief W: Evaluation of a behavioral-medical inpatient rehabilitation treatment including booster sessions: a randomized controlled study. The Clinical Journal of Pain 2009; 25: 356-64
		Sohr G: Effekt einer strukturierten Vorgabe bei der sozialmedizinischen Beurteilung auf die Zufriedenheit, das Empfinden von Fairness, die Depressivität und das Schmerzempfinden bei chronischen Schmerzpatienten. Dissertation, Philipps Universität Marburg; 2008
24	Stier 2001	Stier R, Gerdes N, Bührlen B, Haaf HG, Jäckel WH: "Rekonditionierung in Gruppen" - Ein wirksames Rehabilitationsprogramm bei chronischen Rückenschmerzen? Rehabilitation 2001; 40:321-331
25	Streibelt 2006	Streibelt M, Bethge M: Effects of intensified work-related multidisciplinary rehabilitation on occupational participation: a randomized-controlled trial in patients with chronic musculoskeletal disorders. Int. J. Rehabil Res 2014; 37: 61-6
		Streibelt M, Hansmeier T, Müller-Fahrnow W: Effekte berufsbezogener Behandlungselemente in der orthopädischen Rehabilitation der Rentenversicherung. Rehabilitation 2006; 45: 161-171
		Streibelt M, Blume C, Thren K, Müller-Fahrnow W: Ökonomische Evaluation einer medizinisch-beruflich orientierten Maßnahme bei Patienten mit muskuloskeletalen Erkrankungen – Eine Kosten-Nutzen-Analyse aus Rentenversicherungsperspektive Rehabilitation 2008; 47: 150-157
26	Weiß 2013	Weiß J, Quante S, Xue F, Muche R, Reuss-Borst M: Effectiveness and Acceptance of Acupuncture in Patients with Chronic Low Back Pain: Results of a Prospective, Randomized, Controlled Trial. J Altern Complement Med. 2013; 19: 935-41
27	Fröhlich 2007	Fröhlich SM, Neimeyer R, Greitemann B: RUM-Projekt: Implementierung eines multimodalen Behandlungskonzeptes zur Behandlung chronischer Rückenschmerzen. Orthopädische Praxis 2007; 43: 482-488
		Fröhlich S, Greitemann B, Panning S, Kaiser W: Lebenslust statt Krankheitsfrust. Das evaluierte Rückenfit-Programm der Klinik Münsterland. Praxis Physiotherapie 2012; 3: 148-152
		Greitemann B, Fröhlich S, Neimeyer R. Integriertes orthopädische-psychosomatisches Konzept der Klinik Münsterland. Umsetzung, Replikation und Transfer. RUM-Projekt. Abschlussbericht, Bad Rothenfelde
28	Hüppe 2006	Hüppe A, Glaser-Möller N, Raspe H: Trägerübergreifendes Projekt zur Früherkennung von Rehabilitationsbedarf bei Versicherten mit muskuloskeletalen Beschwerden durch Auswertung von Arbeitsunfähigkeitsdaten: Ergebnisse einer randomisierten, kontrollierten Evaluationsstudie. Gesundheitswesen 2006; 68: 347-356.

Tabelle A2: Aufstellung der ausgeschlossenen Volltexte mit Ausschlussgrund

Ausschlussgrund	Literaturangaben
Beobachtungsstudie ohne KG	Farin E, Glattacker M, Follert P, Kuhl C, Jäckel WH: Effekte und Prädiktoren des Outcomes in der Rehabilitation muskuloskeletaler Erkrankungen. Phys Rehab Kur Med 2004; 14: 123-133
ohne Publikation in Fachjournal	Kluth W: Wirksamkeitsvergleich zwischen verhaltensmedizinisch orientierter orthopädischer Rehabilitation und einer Standardrehabilitation Dissertation, Universität Hamburg 2006
Keine kontrollierte Interventionsstudie	Müller WD, Maier V, Bak P, Smolenski UC: Verzahnung zwischen medizinischer Rehabilitation und beruflich er Reintegration bei Beschäftigten der Automobilindustrie mit Rücken- und Gelenkschmerzen. Phys Rehab Med Kuror 2006; 16:149-154
Durchgeführt in Österreich	Wagner E, Ehrenhofer B, Lackerbauer E, Pawelak U, Siegmeth W: Rehabilitation des chronisch unspezifischen Kreuzschmerzes. Ergebnisse eines multimodalen stationären Behandlungskonzepts. Der Schmerz 2007; 21: 226-233
Review	Karjalainen K, Malmivaara A, van Tulder M, Roine R, Jauhiainen M, Hurri H, Koes B: Multidisciplinary biopsychosocial rehabilitation for subacute low back pain among working age adults. Cochrane Database Syst Rev. 2003;(2):CD002193
Review	Guzmán J, Esmail R, Karjalainen K, Malmivaara A, Irvin E, Bombardier C: Multidisciplinary rehabilitation for chronic low back pain: systematic review. BMJ 2001; 322: 1511-6
Review	Henschke N, Ostelo RW, van Tulder MW, Vlaeyen JW, Morley S, Assendelft WJ, Main CJ: Behavioural treatment for chronic low-back pain. Cochrane Database Syst Rev. 2010 Jul 7;(7):CD002014. doi: 10.1002/14651858.CD002014.pub3.
Review	Waterschoot FP, Dijkstra PU, Hollak N, de Vries HJ, Geertzen JH, Reneman MF: Dose or content? Effectiveness of pain rehabilitation programs for patients with chronic low back pain: A systematic review. Pain 2014; 155: 179-89
Review von Reviews	Smidt N, de Vet HC, Bouter LM, Dekker J, Arendzen JH, de Bie RH, Bierma-Zeinstra SM, Helders PJ, Keus SH, Kwakkel G, Lenssen T et al Exercise Therapy Group: Effectiveness of exercise therapy: a best-evidence summary of systematic reviews. The Australian journal of physiotherapy 2005; 51: 71-85
Review	Rolli Salathé C, Elfering A, Melloh M: Wirksamkeit, Zweckmäßigkeit und Wirtschaftlichkeit des multimodalen Behandlungsansatzes bei chronisch lumbalen Rückenschmerzen. Schmerz 2012; 26:131-149
Review	Morfeld M, Küch D, Greitemann B, Dibbelt S, Salewski C, Franke GH, Liebenau A: Multimodale Interventionsprogramme in der Rehabilitation von Patienten mit chronischen Rückenschmerzen – Ein Vergleich. Rehabilitation 2010; 49: 66-79
Review	Morfeld M, Strahl A, Koch U: Ambulante Rehabilitation in Deutschland. Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz 2011; 54: 420-428
Review	Mau W: Langzeitverlauf und Kostenentwicklung nach Leistungen zur medizinischen Rehabilitation bei muskuloskeletalen Erkrankungen Phys Med Rehab Kuror 2006; 16: 134-138
Review	Hüppe A, Raspe H: Die Wirksamkeit stationärer medizinischer Rehabilitation in Deutschland bei chronischen Rückenschmerzen: eine systematische Literaturübersicht 1980-2001.Rehabilitation 2003; 42: 143-54
Review	Hüppe A, Raspe H: Zur Wirksamkeit von stationärer medizinischer Rehabilitation in Deutschland bei chronischen Rückenschmerzen: Aktualisierung und methodenkritische Diskussion einer Literaturübersicht. Rehabilitation 2005; 44: 24-33
Review	Haaf HG: Ergebnisse zur Wirksamkeit der Rehabilitation. Rehabilitation 2005; 44:e1-e20
Review	Engers A, Jellema P, Wensing M, van der Windt DA, Grol R, van Tulder MW: Individual patient education for low back pain. Cochrane Database Syst. Rev. 2008; 23: CD004057
Review	Bethge M, Müller-Fahnow WM: Wirksamkeit einer intensivierten stationären Rehabilitation bei muskuloskeletalen Erkrankungen: systematischer Review und Meta-Analyse. Rehabilitation 2008; 47: 200 – 209

Tabelle A3 – Übersicht über die Fallzahlen der Studien und die interne Validität

IDN	Kurztitel	Valid N (Netto) letzter MZP	Interne Validität (0-12)	Bias- risiko
01	Bahrke 2006	264	2,00	hoch
02	Bührlen 2002	400	2,00	hoch
03	Bürger 2002	698	1,00	hoch
04	Deck 2012	534	2,00	hoch
05	Dibbelt 2006	481	3,00	hoch
06	Franke 2000	179	4,00	hoch
07	Glattacker 2013	189	2,00	hoch
08	Glattacker 2010	201	3,00	hoch
09	Hampel 2009	153	4,00	hoch
10	Hofreuter 2008	211	4,00	hoch
11	Kainz 2006	872	3,00	hoch
12	Kladny 2003	77	5,00	hoch
13	Kleist 2001	115	6,00	niedrig
14	Klingelhöfer 2003	93* (für Kosten)	3,00	hoch
15	Kolip 2001	225	2,00	hoch
16	Schweikert 2006	264	5,00	hoch
17	Ludwig 2006	243	4,00	hoch
18	Meng 2009	269	7,00	niedrig
19	Mau 2002	413	2,00	hoch
20	Molsberger 2002	124	9,00	niedrig
21	Pomp 2012	361	3,00	hoch
22	Schreiber 2004	284	8,00	niedrig
23	Schwarz 2007	340	8,00	niedrig
24	Stier 2001	131	4,00	hoch
25	Streibelt 2006	102	3,00	hoch
26	Weiß 2013	143	8,00	niedrig
27	Fröhlich 2007	317	3,00	hoch
28	Hüppe 2006	217* (für AU)	5,00	hoch