

## 0.0. Zusammenfassung

**Hintergrund** der Studie ist, daß retrospektive Methoden der Veränderungsmessung nicht in jedem Fall zu Ergebnissen führen, die mit der herkömmlichen Methode – nämlich der indirekten Methode – übereinstimmen (Aseltine et al. 1995; Kohlmann et al. 1998; Schmidt et al. 2000; Bleßmann et al. 2001).

In welcher Weise sich verschiedene Methoden der Veränderungsmessung mit der Erhebung von Parametern der subjektiven Gesundheit zur Dokumentation und Evaluation des Therapieverlaufes und zur Prognose des weiteren Krankheitsverlaufes eignen, ist bisher in Deutschland im Bereich der Rehabilitation nicht untersucht worden.

Daher war Ziel der Studie, zum einen die verschiedenen Methoden der Veränderungsmessung hinsichtlich ihrer Übereinstimmung zu untersuchen, zum anderen herauszuarbeiten, welches die prognostischen Eigenschaften der unterschiedlichen Methoden sind.

Für die folgenden Analysen verwendeten wir drei verschiedene Methoden der Veränderungsmessung: die indirekte Methode und zwei Varianten der direkten Methode. Eine graphische Gegenüberstellung der indirekten und der beiden direkten Methoden ist auf Seite 15 wiedergegeben.

Bei der indirekten Methode handelt es sich um eine Erhebung, bei der die interessierenden Merkmale zu definierten Beobachtungszeitpunkten – in unserem Falle zu Beginn und am Ende der Rehabilitation – erhoben werden. Die Veränderung errechnet sich dann aus der Differenz zwischen den zum ersten und zweiten Meßzeitpunkt erhobenen Meßwerten.

Bei der ersten Variante der direkten Methode – dem sogenannten „direkten“ Vergleichsurteil – wird der Patient lediglich am Ende der Rehabilitation gebeten, eine wahrgenommene Veränderung gegenüber einem früheren Zustand (in unserem Fall zu Beginn der Rehabilitation) in den Kategorien „verbessert – unverändert – verschlechtert“ zu beschreiben (z.B.: *Wenn Sie jetzt noch einmal an Ihre Schmerzen in den 7 Tagen vor der Rehabilitation denken – haben sich Ihre Schmerzen durch die Rehabilitation verändert?*)

Die zweite Variante der Veränderungsmessung wird als „erinnertes Prä“ bezeichnet. Hier wird der Patient am Ende der Rehabilitation aufgefordert, neben der aktuellen Einschätzung seiner gesundheitlichen Situation diese zu einem früheren Zeitpunkt einzuschätzen und zwar anhand derselben Items und Skalen, wie sie zur Erhebung des aktuellen Zustandes verwendet wurden. In diesem Falle errechnet sich die Veränderung aus der Differenz zwischen „erinnertem Prä“ und „Post-Messung“ (erhobener Meßwert am Ende der Rehabilitation).

**Methodik:** Grundlage der Studie bilden Patienten, die an einer medizinischen Rehabilitationsmaßnahme der Rentenversicherungen teilgenommen haben. Die Stichprobe umfaßt N=426 Patienten, von denen 201 an chronischen Herz-Kreislauf-Erkrankungen und 225 an chronischen muskulo-skelettalen Krankheiten litten. 34% der Befragten sind weiblich; 79% stehen noch im Erwerbsleben. Das mittlere Alter liegt bei 51 Jahren.

An vier Meßzeitpunkten (zu Beginn und am Ende des Reha-Aufenthaltes, ½ Jahr und 1 Jahr nach der Rehabilitation) erfolgte eine standardisierte schriftliche Befragung der Patienten. Neben sozialmedizinischen Erfolgsgrößen (wie z. B. Arbeitsunfähigkeit) wurden verschiedene Aspekte der subjektiven Gesundheit erfaßt, darunter körperliche Funktionseinschränkung, Schmerzintensität, soziale Kontaktfähigkeit.

Zum ersten, dritten und vierten Meßzeitpunkt erhielten alle Patienten zur Erfassung der Outcome-Variablen identische Fragebögen. Zum zweiten Meßzeitpunkt wurden zwei unterschiedliche Fragebogen-Versionen verwendet.

Mittels nach Klinik stratifizierter Randomisierung wurden alle an der Studie teilnehmenden Patienten in zwei Untergruppen aufgeteilt. Die Patienten der Gruppe I wurden am Ende der Rehabilitationsmaßnahme gebeten, neben der aktuellen Einschätzung eine wahrgenommene Veränderung gegenüber einem früheren Zeitpunkt zu beschreiben. Die Patienten der Gruppe II wurden aufgefordert, anhand derselben Skalen und Items, wie sie für die Erhebung des Post-Status verwendet wurden, ihre Situation zu einem früheren Zeitpunkt einzuschätzen.

Durch dieses Studiendesign war es möglich, in jeder Gruppe jeweils zwei verschiedene Methoden der Veränderungsmessung hinsichtlich ihrer Überein-

stimmung bzw. ihrer prognostischen Eigenschaften zu untersuchen. Von Vorteil war hierbei, daß sich der gruppeninterne Vergleich stets auf dieselben Probanden bezog.

Eine Vergleich der mit beiden direkten Methoden erzielten Veränderungen wurde aufgrund der verschiedenen Gruppenzugehörigkeiten nicht durchgeführt.

Für die Gegenüberstellung der mit den drei Methoden der Veränderungsmessung erzielten Resultate wurden die Differenzscores (Differenz aus Post-Messung und erhobenem Prä-Status und Differenz aus Post-Messung und erinnertem Prä-Status) zu drei Gruppen zusammengefaßt: positive Differenzen wurden als „Verbesserung“, negative Differenzen als „Verschlechterung“ und Differenzen mit dem Wert Null als „unveränderter Zustand“ gewertet.

Beim direkten Vergleichsurteil wurden die fünf Antwortmöglichkeiten zu drei Kategorien zusammengefaßt: die Antworten „stark gebessert“ und „etwas gebessert“ wurden der Kategorie „verbessert“, „stark verschlechtert“ und „etwas verschlechtert“ der Kategorie „verschlechtert“ und die Antwortmöglichkeit „nicht verändert“ der Kategorie „unverändert“ zugeordnet.

Als statistische Maßzahlen zur Beschreibung der Übereinstimmung zwischen den Varianten der Veränderungsmessung wurden der Rankkorrelationskoeffizient  $r(S)$  nach Spearman und die zufallskorrigierte Übereinstimmung nach Cohen berechnet (Kappa, Cohen 1960).

Als exploratives Verfahren zur Untersuchung der prognostischen Eigenschaften der verschiedenen Methoden wurde die multiple lineare bzw. logistische Regression angewendet.

Eine Liste der zu prognostizierenden abhängigen Variablen, die nach einem halben Jahr bzw. einem Jahr erhoben wurden, ist auf den Seiten 79 - 81 dargestellt.

Aus Gründen der Übersichtlichkeit wurden die abhängigen Variablen inhaltlich in Gruppen aufgeteilt. Die erste Gruppe stellt den Zusammenhang zwischen den mit den verschiedenen Methoden ermittelten Veränderungswerten und den abhängigen Variablen, die nach Aspekten der subjektiven Gesundheit fragen, dar. Die zweite Gruppe weist diesen Zusammenhang mit sozialmedizinischen Erfolgsgrößen auf, die dritte Gruppe mit dem Medikamentenkonsum, die vierte Gruppe mit den Variablen, die sich mit den Veränderungen der Arbeitssituation nach der Rehabilitation beschäftigen, und die fünfte und letzte Gruppe zeigt den Zusammenhang

zwischen den einzelnen Methoden und den Variablen, die die Häufigkeit der Arztkontakte und Behandlungen nach der Rehabilitation analysieren.

Für die im folgenden dargestellten Auswertungen wurden pro Dimension bzw. Item jeweils die Fälle herangezogen, die in allen entsprechenden Variablen (Prä-Post-Messung und erinnertes Prä, Prä-Post-Messung und direktes Vergleichsurteil) gültige Werte hatten. Dieses Vorgehen führte zwar bei einigen Variablen zu einem beträchtlichen Verlust an Fällen (bis zu 10%) stellte aber sicher, daß die Vergleiche zwischen den unterschiedlichen Methoden der Veränderungsmessung stets auf derselben Datengrundlage beruhen.

**Ergebnisse:** Die anhand von Rangkorrelationskoeffizienten und dem Kappa-Koeffizienten nach Cohen durchgeführten Analysen zeigten eine mittlere bis mäßige Übereinstimmung zwischen der indirekten und der direkten Methode der Veränderungsmessung unter Verwendung des erinnerten Prä.

Die mittlere Rangkorrelation bzw. der durchschnittliche Übereinstimmungskoeffizient Kappa betragen 0.50 bzw. 0.42; die Werte der Koeffizienten liegen in einem Bereich von 0.30 bis 0.68 ( $r(S)$ ) und 0.20 bis 0.63 (Kappa). Ungünstigere Ergebnisse zeigten sich im Vergleich der indirekten Methode der Veränderungsmessung mit dem direkten Vergleichsurteil. Die Werte liegen für  $r(S)$  bzw. Kappa im Durchschnitt bei 0.22 [Min: 0.04; Max: 0.3] bzw. bei 0.20 [Min: 0.08 und Max: 0.36]. Hier ist mit einer nur schlechten Übereinstimmung zwischen den mit beiden Methoden ermittelten Veränderungen auszugehen (S.75-76).

Geht man davon aus, daß es sich bei den mit der Prä-Post-Messung ermittelten Veränderungen um Referenzwerte handelt und vergleicht sie mit den Veränderungen, die mit beiden direkten Methoden ermittelt wurden, so fällt auf, daß beide direkten Methoden den Therapieerfolg sowohl unter- als auch überschätzen.

Anhand der erklärten Varianzen ( $R^2$  in der multiplen linearen Regression), mit denen die Prädiktion der indirekten Methode versus der erinnerten Prä-Methode bzw. der indirekten Methode versus dem direkten Vergleichsurteil untersucht wurde, zeigten sich nach einem halben Jahr keine systematischen Unterschiede in der Prognosefähigkeit der indirekten versus der beiden direkten Methoden der Veränderungsmessung.

Es fällt jedoch auf, daß im Vergleich „indirekt versus direkt“ in nahezu allen Variablen-Gruppen mit der indirekten Methode die entsprechenden abhängigen Variablen besser prädiziert werden als mit dem direkten Vergleichsurteil (siehe Tab. 4.3.4.1/S.93). Ein umgekehrtes Bild ergibt sich für die Vergleichsgruppe „indirekt versus erinnertes Prä“. Hier wird außer in der Gruppe I, in der es um Aspekte der subjektiven Gesundheit geht, mit dem erinnerten Prä die meisten abhängigen Variablen besser prädiziert als mit der indirekten Methode.

Ein Jahr nach der Rehabilitation haben sich die prognostischen Eigenschaften der einzelnen Methoden verändert.

In Gruppe I und II (subjektive Gesundheit, sozialmedizinische Erfolgsgrößen) zeigt sich ein systematischer Unterschied in der Prognosefähigkeit der direkten Methoden gegenüber der indirekten Methode. Beide direkten Methoden scheinen geringfügig besser die abhängigen Variablen der entsprechenden Gruppen zu prädizieren als die indirekte Methode (Tab. 4.3.4.2/S.94).

Vergleicht man deskriptiv die beiden direkten Methoden untereinander, so scheint in der Gruppe I der subjektiven Gesundheit das direkte Vergleichsurteil ansatzweise besser zu prädizieren als das erinnerte Prä. In Gruppe II (sozialmedizinische Erfolgsgrößen) hingegen ist kein Unterschied zwischen den Methoden zu erkennen.

Betrachtet man die erzielten Ergebnisse in den Variablen-Gruppen III – V, so sind keine deutlichen Unterschiede mehr zwischen den einzelnen Methoden festzustellen. In Gruppe III scheinen alle untersuchten Methoden der Veränderungsmessung, die abhängigen Variablen gleich schlecht zu prädizieren. Auch die mit der logistischen Regression ergänzend durchgeführten Untersuchungen in den Gruppen IV und V zeigen keine eindeutigen Unterschiede in der Prognosefähigkeit der einzelnen Methoden (S.95).

**Diskussion:** Anhand der in einer großen Stichprobe unter Einschluß zweier rehabilitationsmedizinisch relevanter Erkrankungsgruppen erzielten Ergebnisse kann zusammenfassend festgehalten werden, daß die indirekte und beide Varianten der direkten Methode anhand üblicher Kriterien nur schlecht bis mäßig übereinstimmen. Die Übereinstimmung zwischen der indirekten und der auf dem direkten Vergleichsurteil beruhenden direkten Methode fiel besonders niedrig aus.

Ein systematischer Unterschied in der Prognosefähigkeit der indirekten versus direkten Methoden der Veränderungsmessung ist nach einem halben Jahr nicht zu erkennen.

Ein Jahr nach der Rehabilitation ändern sich allerdings die prognostischen Eigenschaften der einzelnen Methoden. In den Variablen-Gruppen „subjektive Gesundheit und sozialmedizinische Erfolgsgrößen“ zeigt sich ein systematischer Unterschied in der Prädiktionsfähigkeit der direkten Methoden gegenüber der indirekten Methode. Beide direkten Methoden scheinen geringfügig besser die abhängigen Variablen der entsprechenden Gruppen zu präzisieren als die indirekte Methode (Tab. 4.3.4.2/ S.94).

Vergleicht man die beiden direkten Methoden untereinander, so scheint in der Variablen-Gruppe „subjektive Gesundheit“ das direkte Vergleichsurteil ansatzweise besser zu präzisieren als das erinnerte Prä. In der Variablen-Gruppe „sozialmedizinische Erfolgsgrößen“ hingegen ist kein Unterschied zwischen den Methoden zu erkennen.

Insgesamt betrachtet zeigen sich nur schwache Unterschiede in den prognostischen Eigenschaften der verschiedenen Methoden. Abhängig von den verwendeten Prädiktorvariablen, den zu präzisierenden Outcome-Variablen und dem Eintreten kurz- bzw. langfristiger Ereignisse wird mit den drei untersuchten Methoden besser oder weniger gut präzisiert.

Ein zusätzlicher Informationsgewinn durch eine höhere prognostische Wertigkeit, die den Einsatz der direkten Veränderungsmessung vor der indirekten Methode rechtfertigen könnte, ist nicht eindeutig erkennbar.

Es ist jedoch ansatzweise festzustellen, daß Rehabilitationspatienten ein Jahr nach der Rehabilitation sich in ihrem poststationären Verhalten und Befinden eher an ihren direkt erfragten Wahrnehmungen bzw. an der durch Erinnerungsleistung erzielten Veränderungen orientieren als an den von anderen beobachteten Prä- versus Post-Veränderungen.

Die von uns erzielten Ergebnisse geben keinen besonderen Anlaß, die mehrzeitige indirekte Veränderungsmessung als die Methode der Wahl für die Erhebung und Darstellung von Behandlungsverläufen aufzugeben.

Sie zeigen aber, daß neben der indirekten Veränderungsmessung alternative Methoden – wie die direkten Methoden – sinnvoll angewandt werden können. Die Entscheidung, welche Methode letztendlich in einer Studie verwendet wird, sollte sich nach deren Zeitrahmen richten. Werden für die Studie als Veränderungsmaß retrospektive Messungen in Erwägung gezogen, so sollte der Erhebung des „erinnerten Prä-Status“ der Vorzug vor dem direkten Vergleichsurteil gegeben werden.

## **1.0. Einleitung**

### 1.1. Hintergrund der Untersuchung

Die wissenschaftliche Evaluation medizinischer Maßnahmen hat in den letzten Jahren in nahezu allen Versorgungsbereichen stark an Bedeutung gewonnen.

Die Forderung nach verstärkter Überprüfung der Wirksamkeit und des Nutzens diagnostischer und therapeutischer Leistungen hat insbesondere im Sektor der medizinischen Rehabilitation zur Implementation eines umfassenden Qualitätssicherungsprogrammes geführt, das einen entsprechenden Programmpunkt zur Dokumentation des Rehabilitationsergebnisses enthält.

Dieses Qualitätssicherungsprogramm sowie viele andere Evaluationsstudien dieser Art wurden sowohl mit dem Ziel der Qualitätsentwicklung als auch unter Gesichtspunkten des Versorgungsmanagements und der Versorgungsplanung durchgeführt.

In der Ergebnisevaluation wurden dabei neben sozialmedizinischen und gesundheitsökonomischen Bewertungskriterien und objektiven medizinisch-klinischen Befunden insbesondere solche Zielgrößen betrachtet, die den patientennah erfaßten Parametern der subjektiven Gesundheit bzw. der „gesundheitsbezogenen Lebensqualität“ zugerechnet werden können.

Die Hinwendung zur patientenbezogenen Beschreibung des Gesundheitszustandes ist, so Bullinger (1996), auf drei Entwicklungen zurückzuführen: zum einen auf den Paradigmenwechsel in der Definition von Gesundheit, der in Anlehnung an die WHO-Definition (1991) auch die psychischen und sozialen Komponenten des Gesundheitsbegriffes einbezieht, zum anderen die Veränderung der Bevölkerungsstruktur, die sich zu einem höheren Prozentsatz älterer Personen mit erhöhter Häufigkeit chronischer und langfristig behandlungsbedürftiger Erkrankungen hin entwickelt, deren Effekte auf den Lebenszusammenhang der Patienten jenseits der akuten klinischen Wirkungen noch zu untersuchen sind. Der dritte Aspekt, so Bullinger, betrifft die Skepsis gegenüber der Aussagekraft der bisher klassischen Zielkriterien in der medizinischen Behandlung.

Bei den Parametern der subjektiven Gesundheit handelt es sich in der Regel um eine multidimensional erfaßte Patienten-Selbstbeurteilung der körperlichen und psychischen Verfassung und der sozialen Situation. Diese Kriterien haben sich in vielfältiger Weise als geeignet erwiesen, um gesundheitsbedingte Beeinträchtigungen zuverlässig zu beschreiben und zeitliche Veränderungen aufgrund therapeutischer Interventionen hinreichend genau zu dokumentieren.

Entsprechende psychometrisch geprüfte und im Einsatz praktikable Erhebungsinstrumente sind in ausreichender Zahl und Qualität (auch als deutsche Übersetzung englischsprachiger Originalinstrumente) verfügbar.

Werden derartige Erhebungsinstrumente in der Ergebnisevaluation eingesetzt, steht die Messung von Veränderungen im Zeitverlauf im Vordergrund. Der Erfolg einer Intervention bemißt sich als die absolute oder im Vergleich zu einer Referenzgruppe beurteilte Veränderung zwischen mindestens zwei Bezugszeitpunkten.

Zur praktischen Messung solcher Veränderungen wurden bisher verschiedene Methoden vorgeschlagen und in evaluativen Studien eingesetzt. Hierzu gehören neben der Standardmethode der Veränderungsmessung, bei der die Veränderung von Zielkriterien durch wiederholte Messung dieser Kriterien im Zeitverlauf erfolgt (s.u.), solche Methoden, bei denen die erforderlichen Daten retrospektiv, im Rahmen einer einzigen Messung erhoben werden (s.u.).

Solche retrospektiven Methoden der Veränderungsmessung sind in jüngerer Zeit mehrfach als Alternative zur herkömmlichen Veränderungsmessung mit konkurrierender Erhebung der relevanten Merkmale vorgeschlagen worden. Auf der Grundlage verschiedener Vergleichsstudien (s.u.) konnte bereits gezeigt werden, daß die retrospektiven Methoden nicht in jedem Fall zu Ergebnissen führen, die mit den nach der herkömmlichen Methode erzielten Ergebnissen übereinstimmen. Es ist jedoch unklar, in welcher Weise sich die beobachteten Unterschiede auf die prognostischen Eigenschaften der Meßmethoden, das heißt auf ihre Fähigkeit, den künftigen Krankheitsverlauf zu präzisieren, auswirken.

Diese Fragestellung ist insbesondere von großem Interesse für die künftige Bewertung des Krankheitsverlaufes chronischer Erkrankungen. Die mit dem Begriff der chronischen Erkrankung verbundene Vorstellung eines sich langsam aber stetig verschlechternden objektiven und subjektiven Gesundheitszustandes entspricht in vielen Fällen nicht der in der klinischen Praxis beobachtbaren Vielfaltigkeit der

wirklichen Krankheitsverläufe. Selbst bei nosologisch einheitlichen Krankheitsbildern findet sich eine zeitliche Dynamik und Variabilität der Verläufe, die es rechtfertigt, eher die Unterschiedlichkeit der Veränderung als deren Konstanz als kennzeichnendes Merkmal chronischer Krankheiten zu werten (Von Korff 1991).

Folglich ist es um so wichtiger, Veränderungen in methodisch angemessener Weise zu erheben.

Auf der Grundlage des bisherigen Kenntnisstandes kann nicht sicher gesagt werden, welche der beschriebenen Methoden besser in der Lage ist, den weiteren Verlauf einer chronischen Krankheit und ihre Folgen zu prognostizieren. Die hinreichend genaue Kenntnis der prognostischen Validität von Erhebungsdaten ist aber für die Evaluation medizinischer Maßnahmen von übergeordneter Bedeutung. Es wäre demnach durchaus denkbar, daß die retrospektiven Methoden trotz ihrer geringeren meßtheoretischen Qualität gegenüber der Standardmethode („indirekte Prä-Post-Messung“) eine vergleichbare, vielleicht sogar eine höhere Fähigkeit besitzen, Informationen über den weiteren Verlauf einer Erkrankung und ihre sozialmedizinischen Folgen zu liefern.

Daher war Ziel der Studie, verschiedene Ansätze der Veränderungsmessung gleichzeitig anzuwenden, um herauszufinden, in welchem Maße sich die Methoden mit konkurrierender und retrospektiver Erhebung von Parametern der subjektiven Gesundheit für die Dokumentation und Evaluation des Therapieerfolges und zur Prognose des weiteren Krankheitsverlaufs eignen.

## 1.2. Methodische Varianten der Veränderungsmessung

Im Umfeld der wissenschaftlichen Ergebnisevaluation kann auf eine Vielzahl unterschiedlicher Meßmethoden (z.B. direkte und indirekte Veränderungsmessung, zielorientierte Ergebnismessung (ZOE), Zielerreichungsskalierung, allgemeine / gruppenbezogene vs. individualisierte Meß- und Bewertungsstrategien), Beurteilungsperspektiven (Patienten, Ärzte, etc.) und Datenarten (z.B. Fragebogendaten, Laboraten, objektive Messungen, Krankenkassendaten) zurückgegriffen werden. Die resultierenden Forschungsdesigns und die notwendigen Datenerhebungen sind folglich in der Regel sehr aufwendig und in der Routineversorgung kaum zu

bewältigen. Erfassungen der Ergebnisqualität im Rahmen der Qualitätssicherung sollen andererseits ökonomisch, möglichst einfach und dennoch aussagekräftig sein (Schmidt et al., 2000).

In der medizinischen Rehabilitation werden zur Feststellung der Ergebnisse von Heilbehandlungen – bezogen auf die Merkmalsbereiche der subjektiven Gesundheit (z.B. gesundheitliches Befinden, psychisches Befinden, körperliche Beeinträchtigungen) – die indirekte und direkte Methode der Veränderungsmessung herangezogen.

Beide Methoden sind mit einem unterschiedlichen Aufwand verbunden. Während man bei der indirekten Methode von einer Mehr-Punkt-Erhebung spricht, wird die direkte Methode als Ein-Punkt-Erhebung durchgeführt. Bei der direkten Methode, welche nur einen Meßzeitpunkt erfordert, spielen Erinnerungsleistungen eine stärkere Rolle als bei der indirekten Methode.

Würden beide Varianten der Ergebnismessung dasselbe Ergebnis erbringen, dann wären die Methoden völlig austauschbar, und man könnte dem „ökonomischsten“ Verfahren den Vorzug geben.

In Abb. 1 (S.15) sind die in unserer Studie untersuchten verschiedenen Methoden der Veränderungsmessung dargestellt. Dabei handelt es sich um die Standardmethode (indirekte Methode) der Veränderungsmessung sowie um zwei Varianten der direkten Veränderungsmessung.

#### 1.2.1. Indirekte Veränderungsmessung

Bevor im Detail auf die Methode der indirekten Veränderungsmessung eingegangen wird, soll kurz etwas zur Namensgebung dieser Methode gesagt werden.

Die sogenannte Standardmethode wird größtenteils als „indirekte“ Veränderungsmessung bezeichnet. Diese Bezeichnung kann leicht mißverstanden werden, zumal es sich bei dieser Methode um eine in Serie geschaltete „direkte“ Erhebung von Merkmalen handelt. Das heißt, die interessierenden Zielgrößen werden zu mehreren aufeinanderfolgenden Meßzeitpunkten erhoben. Daher sollte man vielmehr von einer „seriellen“ Methode sprechen.

Die Verwirrung wird erst recht perfekt, wenn man diese Methode mit der „direkten“ Methode der Veränderungsmessung vergleicht. Bei dieser Methode wird nämlich ein Merkmal retrospektiv aus der Erinnerung heraus gemessen und nicht - wie man aus ihrer Bezeichnung folgern könnte - direkt.

Da der Begriff der indirekten Methode in der klassischen Veränderungsmessung stark verankert ist, und es nicht Ziel der Studie war, begriffliche Veränderungen vorzuschlagen und neue einzuführen, soll dieser kleine Exkurs lediglich als Denkanstoß dienen.

Für die Ausführungen in dieser Studie wird einfachheitshalber die ursprüngliche Bezeichnung der indirekten Methode beibehalten.

Wie bereits oben beschrieben werden bei der indirekten Methode der Veränderungsmessung die interessierenden Zielgrößen konkurrent, das heißt zeitgleich zu mehreren, definierten Zeitpunkten (z.B. vor und nach einer medizinischen Maßnahme) erhoben. Die Veränderung läßt sich dann rechnerisch aus der Differenz zwischen den zum ersten und zweiten Meßzeitpunkt erhobenen Meßwerten bestimmen (Abb.1 / S.15). Diese konkurrente Prä-Post-Erhebung wird als das Standardverfahren der Veränderungsmessung bezeichnet (vgl. Kohlmann et al. 1998).

### 1.2.2. Direkte Veränderungsmessung

Aus verschiedenen Gründen – besonders aber wegen des mit einer mindestens zweimaligen Messung verbundenen logistischen Aufwandes – sind verschiedene Alternativen zum Standardverfahren der Prä-Post-Messung (indirekte Methode) vorgeschlagen worden. Gemeinsam ist diesen Alternativen, daß sie den Versuch unternehmen, Veränderungen durch retrospektive Erhebungen an nur einem Zeitpunkt abzubilden.

#### 1.2.2.1. Direktes Vergleichsurteil (Variante A der direkten Methode)

Besonderes Interesse hat in der Vergangenheit das direkte Vergleichsurteil der Veränderungsmessung (Variante A der direkten Methode) gefunden, bei der Veränderungen nicht in der Form eines Vergleichs von konkurrent erhobenen Zustandsaussagen (siehe „indirekte Veränderungsmessung“), sondern als retrospektive Vergleichsaussagen operationalisiert werden (vgl. Schmidt et al. 1994). Die Probanden werden in diesem Fall gebeten, eine wahrgenommene Veränderung gegenüber einem früheren Zeitpunkt in entsprechenden Kategorien (z.B. verbessert – unverändert – verschlechtert) zu beschreiben (Abb.1 / S.15).

#### 1.2.2.2. Erinnertertes Prä (Variante B der direkten Methode)

Eine weitere Variante der direkten Veränderungsmessung ist die mit einer konkurrenten Erhebung des Post-Status verbundene retrospektive Messung des früheren Gesundheitszustandes (z.B. vor der Rehabilitationsmaßnahme). In diesem Fall werden die Befragten aufgefordert, anhand derselben Skalen und Items, wie sie für die Erhebung des Post-Status verwendet werden, ihre Situation zu einem früheren Zeitpunkt einzuschätzen. Wie in der indirekten Veränderungsmessung können dadurch Veränderungen aus den Unterschieden zwischen der retrospektiven Prä- und der konkurrenten Post-Messung bestimmt werden.

Diese Methode wird wegen ihres Berechnungsweges auch als pseudo-indirekte oder quasi-indirekte Methode der Veränderungsmessung bezeichnet. Da sie allerdings zu den Ein-Punkt-Erhebungen gehört, das heißt, durch eine einmalige Befragung kann eine Veränderung direkt durch Differenzbildung ermittelt werden, wird sie in dieser Studie als zweite Variante der direkten Methode der Veränderungsmessung zugeordnet.

**Abb. 1: Darstellung der indirekten und direkten Methoden der Veränderungsmessung**

Beobachtungszeitpunkt  $T_1$                       Beobachtungszeitpunkt  $T_2$

**Indirekte Veränderungsmessung**



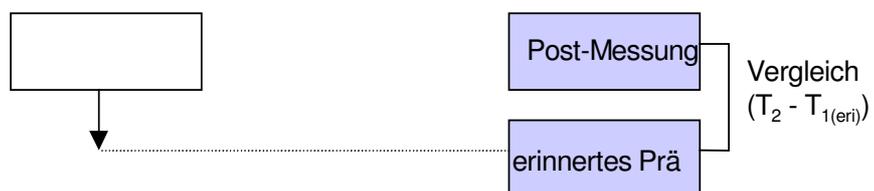
Bei der indirekten Veränderungsmessung werden Zustandsangaben konkurrent zu mindestens zwei Meßzeitpunkten erhoben. Die beobachtete Veränderung wird aus der Differenz der beiden Meßwerte bestimmt.

**Direkte Veränderungsmessung (Variante A)  
direktes Vergleichsurteil**



Bei der Variante A der direkten Veränderungsmessung wird retrospektiv die wahrgenommene Veränderung gegenüber einem Bezugszeitpunkt erfaßt.

**Direkte Veränderungsmessung (Variante B)  
erinnertes Prä**



Bei der Variante B der direkten Veränderungsmessung wird retrospektiv der erinnerte Prä-Status erhoben. Die Veränderung wird als Differenz auf dem konkurrent erhobenen Post-Status und dem retrospektiv erhobenen Prä-Status bestimmt.

### 1.3. Übereinstimmung der indirekten und direkten Methoden der Veränderungsmessung

Aus publizierten Vergleichsstudien ist bekannt, daß die mit der indirekten Methode und den direkten Methoden erzielten Ergebnisse nicht hinreichend gut übereinstimmen.

Ziel dieses Kapitels war es, zunächst eine Übersicht über die beobachteten Ergebnisse hinsichtlich der Übereinstimmung zwischen der indirekten und den von uns untersuchten direkten Methoden der Veränderungsmessung zu erstellen.

Dafür wurden die psychologische und die medizinische Datenbank PSYNDEX und MEDLINE nach Schlagworten wie „Veränderung des Gesundheitszustandes – beobachtete Veränderung – wahrgenommene Veränderung – Ergebnisstudie – Erinnerung – Meßmethoden – medizinische Rehabilitation“ durchsucht.

Bei der Literatursuche zeigte sich, daß der Vergleich der verschiedenen Methoden der Veränderungsmessung durch die Jahrzehnte hinweg immer wieder – mal mehr, mal weniger – diskutiert wurde.

Eine direkte Gegenüberstellung aller in unserer Studie untersuchten Methoden der Veränderungsmessung ist nur zweimal in der medizinischen Rehabilitation durchgeführt worden (Kohlmann et al. 1998 und Schmidt et al. 2002). Alle anderen Studien widmen sich dem Vergleich der indirekten Methode und jeweils einer Variante der direkten Methoden der Veränderungsmessung.

Aufgrund der hohen Anzahl an Literaturquellen wurde der Zeitrahmen auf die Jahre 1995 bis 2002 begrenzt.

Im folgenden werden zunächst die Ergebnisse der Studien vorgestellt, die den Vergleich zwischen der **indirekten Methode** und dem **direkten Vergleichsurteil** untersucht haben.

Daran schließen sich die Resultate, die durch den Vergleich der indirekten Methode mit dem „erinnerten Prä“ der direkten Methode aufgedeckt wurden.

Zuletzt werden die beiden Studien vorgestellt, die die ermittelten Veränderungen aller drei vorgestellten Methoden der Veränderungsmessung vergleichend gegenüberstellen.

## **Indirekte Methode versus direktes Vergleichsurteil**

Aseltine et al. haben sich 1995 dem Vergleich der indirekten Methode und dem direkten Vergleichsurteil gewidmet. 800 Frauen, die wegen gynäkologischer Beschwerden zum Teil operiert, zum Teil medikamentös behandelt wurden, und 434 Männer, die wegen einer benignen Prostatahyperplasie operiert wurden, wurden vor und drei Monate nach der Behandlung gebeten, ihren allgemeinen Gesundheitszustand sowie ihre Krankheitssymptome jeweils mit einer 5-Punkte Skala (schlecht – exzellent) einzuschätzen.

Zum zweiten Befragungszeitpunkt wurden sie zusätzlich gebeten, eine wahrgenommene Veränderung gegenüber ihrem Gesundheitszustand vor der Behandlung in den Kategorien sehr verbessert – etwas verbessert – unverändert – etwas verschlechtert – sehr verschlechtert zu beschreiben.

Die Autoren stellten fest, daß die mittels der indirekten Methode und dem direkten Vergleichsurteil erzielten Veränderungen nicht übereinstimmten. Insbesondere fiel ihnen auf, daß mit dem direkten Vergleichsurteil in der Regel der Behandlungserfolg überschätzt wurde. Bei einem Großteil der Probanden wurde der Gesundheitszustand retrospektiv als verbessert bewertet, obwohl sich ihr Zustand prospektiv gesehen verschlechtert hatte. Dieses Phänomen schien - so Aseltine et al. - bei der Beurteilung des allgemeinen Gesundheitsstatus deutlicher aufzutreten als bei der Bewertung der Krankheitssymptome. Auch die Behandlungsmaßnahme hatte entscheidenden Einfluß darauf, inwieweit die mit den verschiedenen Methoden erzielten Veränderungen voneinander abwichen. So ähnelten die mit dem direkten Vergleichsurteil erzielten Ergebnisse bei den medikamentös behandelten Patienten eher den Ergebnissen der indirekten Methode, während dies bei den operierten Patienten nicht der Fall war.

Aseltine et al. schlossen daraus, daß chirurgische Patienten anscheinend einem derartigen inneren Streß bzw. Druck unterliegen, den operativen Eingriff vor sich rechtfertigen zu müssen, daß sie ihren Gesundheitszustand in der Regel, auch wenn er sich nach der Operation nicht unbedingt verbessert hat, trotzdem als verbessert einschätzen. Zuletzt erwähnten die Autoren, daß retrospektive Beurteilungen in relevantem Umfang vom Zustand zum zweiten Befragungszeitpunkt abhängen. Wenn die Patienten retrospektiv gefragt wurden, ihren Gesundheitszustand vor der Behandlung mit dem nach der Behandlung zu vergleichen, dann schienen sie vielmehr die Frage zu beantworten „Wie gesund

fühlen sie sich jetzt?“ und neigten dazu, ihren tatsächlichen Gesundheitszustand vor dem Eingriff zu vernachlässigen.

Ein gleiches Phänomen beobachteten Kastner et al. (1997) und Raspe et al. (1998). Kastner et al. griffen beispielsweise die Frage auf, ob Veränderungsfragebögen wirklich Veränderungen messen. Dafür verglichen sie die mittels der indirekten Methode und der direkten Methode erzielten Ergebnisse miteinander. Es sollte geklärt werden, ob subjektive Erfolgsbeurteilungen zur Veränderungsmessung geeignet sind und ob damit eine Alternative zu den aufwendigeren Mehr-Punkt-Messungen geschaffen werden kann.

In die Studie wurden 82 Probanden einbezogen, die an einem psychologischen Gruppenprogramm für chronische Kopf- und Rückenschmerzen teilnahmen. Neben klinischen Fragebögen zu Beschwerden, Stimmung und Lebensqualität sowie Schmerztagebüchern vor und nach der Therapie (Instrumente der indirekten Veränderungsmessung) bearbeiteten die Teilnehmer einen 15 Item umfassenden „Fragebogen zur subjektiven Erfolgsbeurteilung der Therapie“ (FSET) nach Therapieende. Zur Überprüfung des Zusammenhanges zwischen Therapieerfolgsmaß und dem FSET wurden Korrelationen durchgeführt. Insgesamt fiel die Übereinstimmung zwischen der Erfolgsbeurteilung durch den FSET und den zu den zwei Meßzeitpunkten erhobenen Variablen und den daraus errechneten Differenzen gering aus.

Es zeichnete sich vielmehr ab, daß der FSET eher als wahrgenommene Veränderung das subjektive Befinden der Patienten nach Beendigung der Therapie widerspiegelte.

Ergänzend wurden Produkt-Moment-Korrelationen zwischen dem FSET und den Meßwerten der Patienten in den klinischen Fragebögen und Schmerztagebüchern vor Beginn und nach Beendigung der Therapie berechnet. Es konnte bestätigt werden, daß die Zusammenhänge der subjektiven Erfolgsbeurteilung mit den nach der Therapie gemessenen Werten deutlich größer waren als mit den vor der Therapie erhobenen. Das heißt, günstige Werte am Ende der Therapie führten zu einer guten Erfolgsbeurteilung.

Obwohl der FSET entgegen der Frageformulierung nicht die durch die Therapie erzielten Veränderungen mißt, sondern eher das Befinden der Patienten nach der Therapie abbildet, plädieren Kastner et al. trotzdem für den weiteren Einsatz dieses direkten Veränderungsmeßinstrumentes in der klinischen Praxis. Dadurch erhält der

Therapeut – so Kastner et al. – in ökonomischer Weise eine wertvolle Rückmeldung über die Zufriedenheit der Patienten mit den durchgeführten Maßnahmen.

Raspe et al. (1998) betrachteten an einer konsekutiven Stichprobe von N=610 Patienten mit muskulo-skelettalen, cardio-pulmonalen und weiteren Erkrankungen die Korrelation der direkten Veränderungsmessung (direktes Vergleichsurteil) mit dem poststationären Wert und der Prä-Post-Differenz (indirekte Methode).

Dafür wurden 2 Wochen vor und 4 bis 6 Wochen nach einer stationären Rehabilitationsmaßnahme die Patienten postalisch befragt. Der poststationäre Fragebogen enthielt neben den Items zur Erhebung der aktuellen Situation Items, mit denen die Patienten eine subjektiv wahrgenommene Veränderung gegenüber der prä-stationären Ausgangslage angeben sollten. Raspe et al. konnten zeigen, daß die retrospektiven Beurteilungen deutlich mit dem Post-Status korrelierten. Er folgerte, daß mit einer unerwünschten „Kontamination“ der Veränderungsbeurteilung durch den zum Befragungszeitpunkt bestehenden Gesundheitszustand zu rechnen sei. Das heißt, daß direkte Vergleichsurteil bildet eher die poststationäre Situation ab als die zwischen den beiden Meßzeitpunkten erfolgte Veränderung.

Nach Abwägen der Vor- und Nachteile beider Methoden setzt sich Raspe vornehmlich für die indirekte Veränderungsmessung ein, die seines Erachtens die wahrscheinlich einzige Methode ist, die eine angemessene Auswertung von Verlaufsdaten ermöglicht. Allerdings weist er auf die noch offene Frage hin, welche der beiden Methoden die höhere prognostische Wertigkeit besitzt.

Es könnte sein, daß sich Rehabilitationspatienten in ihrem poststationären Verhalten und Befinden eher an ihren direkt erfragbaren Wahrnehmungen orientieren als an den von anderen beobachteten Prä- vs. Post-Veränderungen. Solange dies nicht geklärt ist, neigt er dazu, für die Beschreibung chronischer Krankheitsverläufe der indirekten Methode den Vorzug zu geben.

Feine et al. (1998) kamen zu dem Schluß, daß von Patienten selbst wahrgenommene Verbesserungen hinsichtlich chronischer Schmerzen nicht unbedingt den tatsächlichen Schmerzveränderungen entsprechen, sondern vielmehr von der Erinnerungsleistung, dem entsprechenden Befragungszeitraum, der Schmerzintensität vor der Behandlung und zum zweiten Befragungszeitpunkt abhängen.

Obwohl Haas et al. (2002) in einigen Punkten nicht mit den Ergebnissen von Feine et al. übereinstimmte, stellte auch er fest, daß eine durch den Patienten wahrgenommene Verbesserung des Schmerzzustandes innerhalb eines Jahres zu gering mit der aktuellen Schmerzstiuation in Verbindung steht, und diese Meßmethode somit ein ungenaues Instrument zur Einschätzung der klinischen Situation von Patienten darstellt.

Die Autoren folgerten daraus, daß ein Arzt objektive Daten benötigt, um dem Patienten realistisch die aktuelle Veränderung seiner Schmerzsituation zu vermitteln.

Während die bisher vorgestellten Studien eine geringe Übereinstimmung der mit der indirekten Methode und dem direkten Vergleichsurteil erzielten Ergebnisse zeigten, wurde diese Feststellung von den zwei folgenden Studien relativiert.

Mit dem Hintergrund, die Sensitivität des SF-36 Health Survey, nämlich gesundheitliche Veränderungen innerhalb eines Jahres (1992 – 1993) an einer jungen weitestgehend gesunden Population zu überprüfen, untersuchte Perneger (1997) im Rahmen einer prospektiven Studie in Genf den Zusammenhang zwischen den mit der indirekten Methode und dem direkten Vergleichsurteil erzielten Veränderungen unter Anwendung des SF-36.

Der SF-36 beinhaltet 35 Items, mit denen 8 Dimensionen von Gesundheit gemessen werden und ein singuläres Item, mit dem nach einer wahrgenommenen Veränderung des allgemeinen Gesundheitszustandes gegenüber dem vor einem Jahr gefragt wird. Für dieses Item stehen fünf Antwortmöglichkeiten von sehr verbessert bis sehr verschlechtert zur Verfügung.

Die mittels der indirekten Methode für jede Dimension erzielten Differenzscores wurden mit dem Ergebnis des Einzelitems verglichen. Dabei konnte folgendes beobachtet werden: Während der Studienzeit von einem Jahr wurde mit beiden Methoden im Durchschnitt ein stabiler Gesundheitszustand der untersuchten Studienpopulation festgestellt. Die mit der indirekten Methode ermittelten Veränderungen korrelierten gut mit den Antwortkategorien des direkten Vergleichsurteils. Es konnte in sieben von acht Skalen ein signifikanter linearer Zusammenhang zwischen den mit beiden Methoden erzielten Veränderungen beobachtet werden. Eine Ausnahme bildete die Dimension „körperliche Funktionsfähigkeit“. Hier war kein linearer Zusammenhang erkennbar. Die fünf Antwortkategorien des direkten

Vergleichsurteils zeigten eine ordinale Skalierung. Die Intervalle zwischen den einzelnen Antwortmöglichkeiten fielen ungleich aus. So wurden für negative Veränderungen deutlich größere Intervalle ermittelt als für positive Veränderungen. Weiterhin fiel auf, daß die Studienteilnehmer mit dem direkten Vergleichsurteil in der Lage waren, zwischen gesundheitlichen Veränderungen innerhalb verschiedener Zeiträume zu unterscheiden. So korrelierten Veränderungen, die mit der indirekten Methode im Zeitraum 1992 bis 1993 gemessen wurden, nicht mit den wahrgenommenen Veränderungen, die mittels des direkten Vergleichsurteils im Zeitraum 1991 bis 1992 erzielt wurden. Hingegen konnte ein starker Zusammenhang zwischen dem direkten Vergleichsurteil und den Ergebnissen der indirekten Veränderungsmessung im gleichen Befragungszeitraum (1992 – 1993) nachgewiesen werden.

Mit Hilfe des linearen Regressionsmodells wurde die Validität der indirekten Methode, nämlich die Antwortvariationen des direkten Vergleichsurteils zu erklären, überprüft. Dabei stellte man fest, daß das direkte Vergleichsurteil besser geeignet sei, Veränderungen im allgemeinen Gesundheitszustand aufzudecken als Veränderungen in körperlichen oder psychischen Gesundheitsbereichen. Perneger et al. kamen zu dem Schluß, daß beide Methode einigermäßen sensitiv seien, wahre Veränderungen im Gesundheitszustand zu messen. Sie empfehlen, beide Methoden, wenn immer möglich, in der Veränderungsmessung anzuwenden, da sie genau wie Kastner et al. (1997) der Meinung sind, dadurch wichtige und sich ergänzende Informationen zu erhalten.

Auch Schmidt et al. (2001) konnte im Rahmen der EQUA-Studie - die Ergebnisqualität psychosomatischer Rehabilitation anhand der Zusammenhänge zwischen direktem Vergleichsurteil und indirekter Veränderungsmessung zu untersuchen – eine relativ hohe Korrelation zwischen den mit beiden Methoden erzielten Ergebnissen nachweisen.

Anhand einer konsekutiven Patientenstichprobe von N=461 aus 4 psychosomatischen Fachkliniken sollte die bisherige Feststellung - geringe Übereinstimmung der mittels beider Erfassungsmethoden erzielten Ergebnisse - nochmals überprüft werden.

Dabei wurden exemplarisch 13 Einzelitems des subjektiven gesundheitlichen Befindens, deren Veränderungen parallel indirekt und direkt erfaßt wurden, vergleichend ausgewertet.

Auch diese Untersuchung zeigte, daß die beiden Methoden der Veränderungsmessung funktional nicht völlig äquivalent und austauschbar sind. Mit beiden Methoden wurden mehrheitlich Verbesserungen am Ende der Reha-Maßnahme gemessen, wobei die direkte Methode insgesamt ein etwas „günstigeres“ Bild der Ergebnisseite vermittelte. Die Höhe der Korrelation zwischen den Methoden auf Skalen-Niveau bei  $r(S)=0.65$  demonstrierte allerdings, daß der wahre Zusammenhang trotz anderweitiger Aussagen beachtlich hoch ausfiel. Das würde bedeuten, daß in beide Veränderungsmaße vergleichbare Aspekte eingingen. Eine genauere Analyse, die sich auf die Schnittmenge dieser gemeinsamen Veränderungsaspekte konzentriert, soll demnächst – so Schmidt et al. – durchgeführt werden.

Neben dem Vergleich der indirekten Methode versus dem direkten Vergleichsurteil haben sich zahlreiche Studien mit dem Vergleich der **indirekten Methode** und der zweiten Variante der direkten Veränderungsmessung (der **erinnerten Prä-Methode**) beschäftigt.

Auch hier wurden für den Literaturüberblick von 1995 bis 2002 die Datenbanken PSYINDEX und MEDLINE nach Schlagworten wie „Veränderung des Gesundheitszustandes – beobachtete Veränderung – Ergebnisstudie – Erinnerung – Meßmethoden – medizinische Rehabilitation“ durchsucht.

### **Indirekte Methode versus erinnerte Prä-Methode**

Mancuso et al. widmete sich 1995 dem Vergleich der indirekten und der erinnerten Prä-Methode (zweite Variante der direkten Veränderungsmessung). Beide Methoden wurden zur Beurteilung der therapeutischen Wirksamkeit von medizinischen und chirurgischen Behandlungen angewandt.

Sie verfolgten das Ziel, die Auswirkung eines möglicherweise fehlerhaften Erinnerungsvermögens beim Vergleich des faktisch beobachteten Gesundheitszustandes vor einer Hüftoperation (TEP<sup>1</sup>) mit der retrospektiv einige Jahre nach der Operation durch Erinnerungsleistung erfaßten präoperativen Situation aufzudecken. 104 Patienten bekamen vor einer geplanten Hüftoperation den sog. „Hip Rating Questionnaire (HRQ)“, ein krankheitsspezifischer Fragebogen, der die Hüftsymptome

---

<sup>1</sup> TEP= Totalendoprothese

und die Auswirkung der Hüfterkrankung auf die Lebensqualität des Patienten in vier Dimensionen (Schmerz, Laufvermögen, Funktionsfähigkeit im Alltag und Auswirkung der Hüftarthritis auf den Allgemeinzustand) mißt. Einige Jahre nach der Hüftoperation (im Mittel 2,5 Jahre) wurden dieselben Patienten gebeten, nochmals den HRQ-Fragebogen auszufüllen. Parallel dazu wurden sie telefonisch kontaktiert und aufgefordert, retrospektiv ihren Zustand vor der Hüftoperation anhand des HRQ-Fragebogens zu beschreiben.

Als statistische Maßzahlen zur Beschreibung der Übereinstimmung zwischen dem „aktuell“ erhobenen präoperativen Zustand und dem „erinnerten“ Zustand wurden die zufallskorrigierte Übereinstimmung nach Cohen und der Rangkorrelationskoeffizient  $r(S)$  verwendet. In drei Dimensionen (Schmerz, Funktionsfähigkeit und Auswirkung der Arthritis auf den Allgemeinzustand) zeigte sich eine schlechte bis mäßige Übereinstimmung (Kappa von 0.13 bis 0.42); in der vierten Dimension (Laufvermögen) konnte eine moderate Übereinstimmung (Kappa 0.45 – 0.57) festgestellt werden.

Anscheinend haben Patienten, die sich einer Hüftoperation unterzogen haben, nach ca. 2 ½ Jahren Probleme, ihren präoperativen Zustand in korrekter Weise zu beschreiben. Es zeigte sich, daß die Patienten postoperativ stärkere Schmerzen, besseres Laufvermögen, bessere Funktionsfähigkeit im Alltag und eine schlimmere Beeinträchtigung durch die Erkrankung im Leben vor der Operation angaben, als sie es damals präoperativ beschrieben hatten. Um herauszufinden, ob es sich bei der Erinnerungsleistung eventuell um einen systematischen Fehler (Bias) handelte, wurden die Patienten in Untergruppen stratifiziert und der McNemar Test durchgeführt. Ein systematischer Fehler konnte nicht nachgewiesen werden, da sich zum einen die Prozentangaben der Patienten, die einem Erinnerungsfehler unterlagen, in den einzelnen Subgruppen unterschieden, und zum anderen deutliche Diskrepanzen in der Größe und der Richtung des Fehlers auftraten. Der Vergleich der mit beiden Methoden ermittelten Veränderungen zeigte, daß mit der erinnerte Prä-Methode der Behandlungserfolg nach einer Hüftoperation von 68% aller Patienten überschätzt wurde.

Mancuso et al. schlossen daraus, daß die erinnerte Prä-Methode keine valide Methode ist, einen präoperativen Zustand retrospektiv zu beschreiben. Vielmehr kann sie zu einer Über- bzw. Unterschätzung eines Behandlungserfolges führen.

Everts et al. (1999) fanden heraus, daß Patienten mit drohendem oder tatsächlich eingetretenem Herzinfarkt dazu neigen, 6 Monate nach dem akuten Ereignis, die Intensität der damals empfundenen Brustschmerzen zu überschätzen.

Dafür wurden 177 Patienten zu Hause oder in den ersten Stunden im Krankenhaus gebeten, die Stärke der aktuellen Brustschmerzen anhand einer numerischen Skala einzuschätzen. Sechs Monate nach dem Vorfall wurden sie aufgefordert, die damals empfundene Schmerzintensität anhand derselben Skala, wie sie zur Erhebung des früheren Zustandes verwendet wurde, anzugeben.

Beim Vergleich der Mittelwerte unter Berücksichtigung verschiedener demographischer und subjektiver Charakteristika (Angst, emotionaler Zustand, Zufriedenheit, kardiale Symptome) konnte eine systematische Überschätzung der Brustschmerzintensität festgestellt werden, die von der vom Patienten damals empfundenen Angst abhängig war.

Lingard et al. (2001), die sich ebenfalls mit dem Vergleich zwischen indirekter und direkter Methode unter Verwendung des erinnerten Prä beschäftigten hatten, zeigten, daß der nach drei Monaten retrospektiv erhobene präoperative Schmerz- und Funktionszustand von Patienten nach Einsatz eines künstlichen Kniegelenkes nur mäßig mit der vor der Operation angegebenen Situation übereinstimmte (Kappa 0.20 – 0.41). Sie ermahnten, daß bei der Erhebung des präoperativen Zustandes mittels der retrospektiven Methode deren Grenzen unbedingt berücksichtigt werden sollten, bevor auf die Effektivität derartiger Operationen geschlossen würde.

Auch die Untersuchung von Dawson et al. (2002) bestätigte die Ergebnisse der vorherigen Studien. Wie Mancuso und Lingard überprüften sie die statistische Übereinstimmung von prospektiv und retrospektiv erhobenen Daten mittels Cohen's Kappa.

Dazu wurden Patientenangaben zum Schmerzzustand zu Beginn der „National Low back Pain Study“ in Amerika mit retrospektiv erhobenen Daten – zehn Jahre nach Beginn der Studie wurden die Patienten telefonisch gebeten, ihre damalige Schmerzsituation nochmals einzuschätzen – verglichen.

Insgesamt zeigte sich eine mäßige Übereinstimmung zwischen den initial und nach zehn Jahren erhobenen Angaben zur Schmerzsituation (Kappa 0.37).

Es konnte allerdings festgestellt werden, daß bestimmte Informationen über den Schmerz auch nach zehn Jahren genauer angegeben werden konnten als andere. Beispielsweise erzielten Fragen zu Schmerzhäufigkeit oder schmerzauslösende

Bewegungen bessere Übereinstimmungswerte zwischen prospektiv und retrospektiv erhobenen Angaben als Fragen zur damals empfundenen Schmerzintensität.

Dawson et al. folgerten daraus, daß nur einige Fragen sinnvoll zur Erhebung retrospektiver Daten eingesetzt werden sollten, um der potentiellen Gefahr, Behandlungserfolge zu überschätzen, zu begegnen.

Wie auch schon beim Vergleich zwischen indirekter Methode und dem direkten Vergleichsurteil gezeigt, unterscheiden sich die in den verschiedenen Studien erzielten Ergebnisse deutlich voneinander. Die bisher dargestellten Studien vertreten den Standpunkt, daß die indirekte Methode und die „erinnerte Prä-Methode“ nur mäßig in ihren Ergebnissen übereinstimmen. Die im folgenden dargestellten Untersuchungen hingegen zeigen, daß man durchaus von einer zufriedenstellenden Übereinstimmung beider Methoden ausgehen kann.

Stallmann (1999) und Steffanowski et al. (2002) haben sich mit der Güte von retrospektiven Angaben von Patienten beschäftigt.

Stallmann stellte Selbstangaben der Patienten aus einer standardisierten Befragung zur Inanspruchnahme von Leistungen zur medizinischen Rehabilitation Daten aus der Dokumentation der Rentenversicherung gegenüber. Dabei sollte geklärt werden, inwieweit eine Übereinstimmung von Eigenangaben und Angaben der Rentenversicherung bestehen. Die konkrete Frageformulierung nach der Inanspruchnahme von Rehabilitationsleistungen bezog sich zum einen ganz allgemein darauf, ob die Befragten überhaupt schon einmal an einer stationären Heilbehandlung in einer Kur- oder Rehabilitationsklinik teilgenommen hatten, und wenn ja, wie häufig dies der Fall gewesen war. Zum anderen sollte für den letzten Aufenthalt das Jahresdatum, der gesundheitliche Grund und der Kostenträger angegeben werden.

Es zeigte sich insgesamt ein Übereinstimmungsanteil von 89%, allerdings war dies im Fall selbstberichteter Inanspruchnahme nur zu 63% deckungsgleich mit den Rentenversicherungsangaben, während sich bei der Nichtinanspruchnahme ein Anteil von 99% ergab. Alle Angaben wurden zusätzlich unter Berücksichtigung verschiedener soziodemographischer Gesichtspunkte (Alter, Geschlecht, Berufsstatus) vergleichend untersucht.

Folgende Zusammenhänge konnten aufgedeckt werden: Eine Negativauskunft auf die Frage nach früheren Rehabilitationsmaßnahmen ist in den meisten Fällen sehr zuverlässig gewesen. Wurden frühere Rehabilitationsmaßnahmen in Anspruch

genommen und sollten diese differenziert nach Kostenträgerschaft erfaßt werden, so waren die Antworten zuverlässiger, wenn nur eine Rehabilitationsmaßnahme in der Vergangenheit aufgetreten war, die von der Rentenversicherung bezahlt wurde. Sobald mehrere unterschiedliche Ereignisse oder Behandlungsmaßnahmen in der Vergangenheit aufgetreten waren, stieg die Wahrscheinlichkeit unzuverlässiger Angaben. Je weiter ein Ereignis zurücklag, um so schwieriger wurde die genaue Erinnerung daran. Die soziodemographische Differenzierung der Daten wies auf die Bedeutung des Alters hin. Jüngeren Patienten fiel die Antwort auf die Frage nach Rehabilitationsmaßnahmen leicht, da sie besonders häufig mit „nein“ antworten konnten, und dies dann auch mit den Angaben der Rentenversicherung übereinstimmte. Bei den Älteren hingegen war es wahrscheinlicher, daß sie bereits Rehabilitationsmaßnahmen in Anspruch genommen hatten. Diese mußten sie dann einem Kostenträger zuordnen, womit die Möglichkeit inkonsistenter Antworten stieg. Unterschiede im zuverlässigem Antwortverhalten zwischen verschiedenen Berufsgruppen konnten nicht entdeckt werden. Das Merkmal Geschlecht hatte nur wenig Bedeutung für die Konsistenz der Angaben.

Insgesamt beurteilte Stallmann die Güte der Eigenangaben der Patienten zur medizinischen Rehabilitation als sehr zufriedenstellend.

Steffanowski et al. (2002) setzte im Rahmen einer multizentrischen Studie zur Erfassung der Ergebnisqualität stationärer psychosomatischer Rehabilitationsbehandlungen (EQUA-Studie) Selbstangaben von Patienten zu Arbeitsunfähigkeitszeiten (AU) im Jahr vor und nach der Behandlung mit parallel erhobenen Krankenkassendaten in Beziehung. Damit verfolgte er das Ziel, die Aussagekraft von Patientenangaben, wenn objektive Krankenkassendaten als Vergleichsmaßstab herangezogen werden, zu überprüfen.

Eine Stichprobe von 183 Patienten wurde zu Beginn der stationären Rehabilitation sowie exakt 12 Monate nach Entlassung aus der Klinik gebeten, die Frage „Wie lang waren Sie in den letzten 12 Monaten insgesamt krank geschrieben?“ zu beantworten. Folgende sieben rangskalierte Antwortmöglichkeiten konnten angekreuzt werden: „nie; weniger als 1 Woche; etwa 1-2 Wochen; etwa 3-4 Wochen; etwa bis 3 Monate; 3-6 Monate; mehr als 6 Monate“. Die Krankenkassendaten wurden mit dem Anfangs- und Enddatum etwaiger AU-Zeiten im Beobachtungszeitraum sowie mit der Dauer der einzelnen AU-Fälle in Tagen erhoben.

Aus Sicht der Patienten und Krankenkassen zeigte sich eine Abnahme der Krankschreibungszeiten nach Abschluß der Reha-Maßnahme. Bezüglich des absoluten Ausmaßes der AU-Zeiten lieferten Patienten- und Kassendaten vergleichbare Werte. Die beobachteten Diskrepanzen waren statistisch nicht signifikant. Der Zusammenhang zwischen Selbst- und Fremdanangaben betrug  $r(S) = 0.78$  ( $p < 0.001$ ) für den Zeitraum 12 Monate vor der Rehabilitation bzw.  $r(S) = 0.61$  ( $p < 0.001$ ) für den Zeitraum 12 Monate nach der Rehabilitation.

Auch Steffanowski beurteilte die rangskalierte Erfassung von AU-Zeiten durch den Patienten, wenn als Vergleichsmaßstab objektive Krankenkassendaten herangezogen werden, als zuverlässig.

Von einer hohen Übereinstimmung zwischen prospektiv und retrospektiv erhobenen Daten sprachen Legler et al. (2000). Sie demonstrierten, daß Patienten, bei denen Prostatakrebs diagnostiziert wurde, durchaus gut in der Lage waren, ein halbes Jahr nach Diagnosestellung ihre damaligen Funktionseinschränkungen einzuschätzen.

133 Männer wurden gebeten, ihre prädiagnostischen Urin-, Sexualfunktion und Verdauungssituation anhand einer Skala von 0 bis 100 einzuschätzen. Nach einem halben Jahr wurden sie aufgefordert, neben der aktuellen Einschätzung ihrer Urin-, Sexual- und Verdauungsfunktionen ihre Situation zum Zeitpunkt der Diagnosestellung erneut zu beschreiben.

Insgesamt konnten Legler et al. eine relativ hohe Übereinstimmung zwischen der beobachteten und der retrospektiv erhobenen Einschätzung der prädiagnostischen Situation feststellen ( $\geq 69\%$ ). Dies – so Legler et al. – widerlege die Meinung, daß Behandlungsmaßnahmen die Erinnerungsfähigkeit der Patienten an ihren prädiagnostischen Zustand beeinflussen.

Auch der Vergleich zwischen den prospektiv und nach 6 Monaten retrospektiv gemessenen Veränderungen wies ein relativ hohes Maß an Konsistenz auf [Korrelationskoeffizienten in den Bereichen Urinfunktion: 0.828; Sexualfunktion: 0.682; Verdauungsfunktion: 0.618], das den zukünftigen Einsatz der retrospektiven Methode in der Veränderungseinschätzung der untersuchten Funktionsbereiche unterstützt.

An dieser Stelle sei aus Gründen der Vollständigkeit noch die Studie von Guadagnoli et al. (1995) erwähnt, der den Effekt der Zeit auf die Einschätzung des prästationären Gesundheitszustandes untersuchte.

Immer häufiger, so Guadagnoli, werden zur Beurteilung der Effektivität von medizinischen Maßnahmen Patienten befragt, eine Veränderung in ihrem Gesundheitszustand durch eine Intervention zu beschreiben. Dies hat in der Regel eine zweimalige Erhebung des Gesundheitszustandes vor und nach einer Maßnahme zur Folge. Die Erhebung des Ausgangszustandes ist allerdings nicht in jedem Falle möglich, beispielsweise bei notfallmäßig aufgenommenen oder initial bewußtlosen Patienten. Daher werden diese Patienten nach der Hilfemaßnahme gebeten, sich an ihren Gesundheitszustand vor der Intervention zurück zu erinnern.

657 Patienten, die wegen eines akuten Herzinfarktes oder zum Ausschluß eines akuten Infarktes ins Krankenhaus eingeliefert wurden, wurden am Ende ihres Aufenthaltes sowie drei Monate danach gebeten ihren Gesundheitszustand vor der Behandlung einzuschätzen. Es sollte überprüft werden, ob bei den Patienten die Erinnerung an den prästationären Zustand zu den verschiedenen Befragungszeitpunkten auf Item-Ebene und auf Skalen-Niveau übereinstimmt.

Er stellte fest, daß die Aussagen auf Skalen-Niveau konsistenter innerhalb der drei monatigen Erinnerungsperiode ausfielen als der individuelle Vergleich der Einzelitems. Im Durchschnitt war nur in zwei Fällen, ein signifikanter Unterschied zwischen den zum ersten und zweiten Befragungszeitpunkt erhobenen Angaben zu verzeichnen. In diesen beiden Fällen schätzten die Patienten drei Monate nach der damaligen Intervention ihre prästationäre körperliche Funktionsfähigkeit etwas besser ein als am Ende des Krankenhausaufenthaltes. Da die durchschnittliche Abweichung weniger als 5 Punkte in einer Skala von 0 bis 100 betrug, folgerte Guadagnoli, daß zu beiden Befragungszeitpunkten - am Ende der Therapie und drei Monate nach der Therapie - der prästationären Gesundheitszustandes ähnlich eingeschätzt wird. Nach seinen Untersuchungen würde ein möglicherweise vorhandener Bias eher dahin tendieren, einen Behandlungserfolg zu unterschätzen.

Leider beschränkte sich die Untersuchung lediglich auf die Erinnerungsleistung der Patienten. Eine objektive Beurteilung des prästationären Zustandes war nicht möglich, so daß keine Aussage darüber gemacht werden konnte, zu welchem Befragungszeitpunkt die prästationäre körperliche Funktionsfähigkeit der Patienten am genauesten reflektiert wurde.

Im folgenden werden die zwei rehabilitationswissenschaftlichen Studien vorgestellt, in denen alle drei Methoden der Veränderungsmessung gleichzeitig angewandt wurden. Dadurch konnten die mit den jeweiligen Methoden erzielten Ergebnisse direkt gegenübergestellt werden.

### **Indirekte Methode versus erinnertes Prä versus direktes Vergleichsurteil**

Kohlmann et al. untersuchte das Thema „Wie gut stimmen „indirekte“ und „direkte“ Methoden der Veränderungsmessung überein?“ Er verwendete ein Studiendesign, mit dem er drei verschiedene Methoden der Veränderungsmessung miteinander vergleichen konnte: die als Differenz aus der prä- und poststationären Messung berechnete „indirekte“ Veränderung, die anhand von erinnerten Prä-Werten bestimmten Veränderungen (Variante B) und die in direkten Vergleichsurteilen (Variante A) erfaßten Veränderungen (vgl. Abb.1/S.15).

Dieses Studiendesign wurde ein Jahr später in meiner Studie angewandt. Vorteil dieses Vorgehens war, daß die erzielten Ergebnisse direkt mit den Ergebnissen von Kohlmann et al. vergleichbar waren.

Seine Untersuchung stützte sich auf die bereits in der Studie von Raspe et al. (1998) erwähnten Daten. Ergänzend muß erwähnt werden, daß der damals verwendete poststationäre Fragebogen neben den Items zur Erhebung der aktuellen Situation und den Items, mit denen die Patienten eine subjektiv wahrgenommene Veränderung gegenüber der prästationären Ausgangslage angeben sollten, Items zur retrospektiven Erhebung ihrer Situation vor dem Rehabilitationsaufenthalt enthielt.

Kohlmann konnte erneut (s.o.) zeigen, daß statistisch eine nur geringe bis mäßige Übereinstimmung der mit den verschiedenen Methoden bestimmten Veränderungen besteht. In 14 Einzelvariablen wurden verschiedene Aspekte der subjektiven Gesundheit erfaßt. Die für diese insgesamt 14 Indikatoren untersuchte zufalls-korrigierte Übereinstimmung (Cohens Kappa) zwischen der indirekten und den Varianten A bzw. B der direkten Veränderungsmessung (Abb. 1/ S.15) lagen im Bereich von 0.07 bis 0.33 bzw. 0.22 bis 0.61.

Weiterhin fiel auf, daß der retrospektiv gemessene Prä-Status sich systematisch und statistisch signifikant von den konkurrent gemessenen Werten unterschied.

Hieraus ergab sich eine Überschätzung der durch die Rehabilitation eingetretenen Verbesserung, wenn als Veränderungsmaß die erinnerte Prä-Methode verwendet wurde. Eine solche Verzerrung konnte er auch für die Variante A (direktes Vergleichsurteil) der direkten Veränderungsmessung nachweisen.

Zur Untersuchung der zwischen den verschiedenen Veränderungsmessungen bestehenden korrelativen Struktur führte er eine Hauptkomponentenanalyse durch. Damit zeigte sich, daß die mit der Methode des direkten Vergleichsurteils erhobenen Aspekte der Veränderung mit hoher Wahrscheinlichkeit eine singulär und eigenständige Meßdimension repräsentierten und nicht der mehrdimensionalen Struktur folgten, die mit den beiden übrigen Methoden darstellbar war.

Die aus der Untersuchung resultierende Empfehlung bestand darin, in rehabilitationswissenschaftlichen Studien die indirekte Methoden der Veränderungsmessung anzuwenden und nur, wenn retrospektive Methoden unumgänglich seien, der Erhebung des „erinnerten Prä-Status“ den Vorzug vor dem direkten Vergleichsurteil zu geben.

Schmidt et al. (2002) ergänzte seine bisherigen Ergebnisse (2001) mit dem zusätzlichen Vergleich zwischen der indirekten und der Quasi-Variante („erinnertes Prä“) der indirekten Veränderungsmessung. Mit der „quasi-indirekten“ Methode werden – wie bereits weiter vorne beschrieben – zeitlich zurückliegende Zustände (z.B. Gesundheitszustand vor Beginn einer Behandlung) retrospektiv („erinnertes Prä“) erfaßt und die interessierenden Veränderungen – analog zu der indirekten Methode – rechnerisch aus den Zustandsangaben bestimmt.

Auch hier sollte die Frage geklärt werden, inwiefern eine Übereinstimmung der mittels der unterschiedlichen Methoden erzielten Ergebnisse vorliegt.

Es wurden exemplarisch drei Merkmale des subjektiven gesundheitlichen Befindens (körperliches Befinden, seelisches Befinden, Gesundheitszustand) herausgegriffen. Zeitgleiche Statusmessungen wurden bei Aufnahme, vor Entlassung und 1 Jahr nach Entlassung durchgeführt. Eine direkte Veränderungseinstufung erfolgte bei Entlassung und 1 Jahr danach. Außerdem wurde 1 Jahr nach Entlassung retrospektiv der Aufnahme- und Entlassungszustand eingeschätzt. Verglichen wurden jeweils die echten und retrospektiven Statureinschätzungen für die Zeitpunkte Aufnahme und Entlassung, die beobachtbaren kurzfristigen und längerfristigen Veränderungen auf der Basis dieser Statusmessungen und die jeweiligen Zusammenhänge mit den direkten Veränderungsmessungen.

Es konnte gezeigt werden, daß echte und retrospektive „Vorher-Nachher-Vergleiche“ zwar Unterschiede in gleicher Richtung anzeigen, jedoch größtenteils nur moderat miteinander korrelierten. Der Vergleich der echten und retrospektiven Statusmessungen für den Aufnahme-Zeitpunkt wies bei allen drei Merkmalen signifikante Unterschiede auf. Der von den Patienten „erinnerte“ Aufnahmestatus fiel in der Regel ungünstiger aus als der reale. Auch die tatsächlichen und retrospektiven Entlassungseinschätzungen unterschieden sich signifikant in allen drei Merkmalen. Hier wurde allerdings der „erinnerte“ Entlassungszustand günstiger eingeschätzt als der tatsächlich beobachtete. Unter Verwendung retrospektiver Statureinschätzungen konnten deutlich höhere Effektstärken bezüglich kurzfristiger und längerfristiger Veränderungen erzielt werden als unter Einbeziehung von tatsächlich beobachteten Statuswerten. Zum Zeitpunkt „1 Jahr nach Entlassung“ korrelierten die aufgrund der retrospektiven Statusmessungen bei Aufnahme gebildeten Differenzwerte durchweg deutlicher mit korrespondierenden direkten Veränderungseinschätzungen als die echten Differenzwerte.

Diesbezüglich verweist Schmidt et al. auf eine interessante Studie von Pratt et al. (2000). Sie liefert Anhaltspunkte dafür, daß herkömmliche Prä-Post-Vergleiche – sofern sich der individuelle Bezugsrahmen für Statureinschätzungen im Laufe der Zeit verändert – eher zu einer Unterschätzung der tatsächlichen Programmeffekte führt. Derartige Effekte ließen sich, so Pratt et al., eventuell durch die retrospektive Prätest-Methodik vermeiden, wenn man davon ausginge, daß Patienten für die Einschätzung der verschiedenen Meßzeitpunkte einen gemeinsamen internen Standard verwenden und somit die Werte zwischen den Zeitpunkten besser zu vergleichen sind (vgl. Schulz et al. 1999).

Schmidt et al. folgerte daraus, daß retrospektive Vorher-Nachher-Vergleiche insgesamt ein „günstigeres“ Bild der Ergebnisse vermitteln als der klassische Ansatz.

Auch Schulz et al. (1999) gab nach Durchführung seiner 1-Jahres Katamnese studie in der psychosomatischen Fachklinik St. Franziska-Stift Bad Kreuznach als möglichen Vorteil der retrospektiven Methoden an, daß Patienten für die Einschätzung verschiedener Meßzeitpunkte einen gemeinsamen internen Standard verwenden und somit die Werte zwischen den Zeitpunkten besser verglichen werden können. Allerdings führte er einschränkend hinzu, daß berücksichtigt werden müßte, ob die Patienten überhaupt in der Lage sind, ausreichend

zuverlässige Einschätzungen für länger zurückliegende Zeitpunkte abgeben zu können, und ob die nachträglichen Einschätzungen nicht auch stark von aktuellen Stimmungen und Ereignissen beeinflusst werden.

Abschließend sind in der folgenden Tabelle (Tab.1) die Vor- und Nachteile der indirekten und direkten Methoden der Veränderungsmessung gegenübergestellt. Unter der Spalte „direkte Methoden“ wurden die Vor- und Nachteile der erinnerten Prä-Methode und des direkten Vergleichsurteils zusammen dargestellt.

**Tab.1:** Gegenüberstellung der Vor- und Nachteile der indirekten und direkten Methoden der Veränderungsmessung

<b>Indirekte Methode</b>		<b>Direkte Methoden</b> „Erinnertes Prä“ und „direktes Vergleichsurteil“	
<b>Vorteil</b>	<b>Nachteil</b>	<b>Vorteil</b>	<b>Nachteil</b>
In der klinischen Forschung als Meßmethode etabliert	Mehr-Punkt-Erhebung (sehr aufwendig)	Ein-Punkt-Erhebung (ökonomisch)	Geringe Übereinstimmung mit indirekter Methode
Mißt zu definierten Zeitpunkten „objektiv“ Werte	Evtl. Wechsel interner Meßstandards, folglich mögliche Unterschätzung des Behandlungserfolgs	Ermöglicht Veränderungsmessung auch dann, wenn Ausgangsstatus nicht erhoben werden konnte	Meist Überschätzung des Behandlungserfolgs
Geringe Abhängigkeit vom poststationären Status		Verwendung eines internen Meßstandards	Mißt eher Zustand zum Befragungszeitpunkt als die Veränderung <sup>2</sup>
			Erinnerungsleistung ist von verschiedensten Faktoren abhängig (Alter, Häufigkeit der Behandlungen, Gemütszustand, etc.)

<sup>2</sup> bezieht sich lediglich auf einen Nachteil des direkten Vergleichsurteils

Eine Entscheidung, welche Methode der Veränderungsmessung letztendlich die „bessere“ oder „geeignere“ für rehabilitationswissenschaftliche Evaluationsstudien ist, kann aus diesen Ergebnissen nicht eindeutig getroffen werden.

Vielmehr zeigt die Zusammenfassung, daß eine Entscheidung, die allein auf dem Vergleich der mit den verschiedenen Methoden ermittelten Veränderungen basiert, Grenzen hat.

Ziel unserer Studie war es daher, zum einen die verschiedenen Methoden der Veränderungsmessung hinsichtlich ihrer Übereinstimmung nochmals zu untersuchen, und zum anderen die prognostischen Eigenschaften der unterschiedlichen Methoden herauszuarbeiten. Dadurch kann möglicherweise ein zusätzlicher Informationsgewinn erzielt werden, der eine abschließende Beurteilung über den Einsatz der entsprechenden Meßmethode in Evaluationsstudien ermöglicht.