

Bachelorarbeit

Die Wirksamkeit von Stabilisationstraining im Vergleich mit manueller Therapie bei chronischen lumbalen Schmerzen

Pascal Oliver Kaufmann

S00-723-908

Allenmoosstrasse 62, 8057 Zürich

Departement:	Gesundheit
Institut:	Institut für Physiotherapie
Studienjahr:	6. Semester, Frühlingsemester 2010
Eingereicht am:	10. September 2010
Betreuende Lehrperson:	Dr. Jan Kool

Inhaltsverzeichnis

1. ABSTRACT	4
2. EINLEITUNG	6
2.1. EINFÜHRUNG IN DIE THEMATIK.....	6
2.2. ZIELSETZUNG.....	7
2.3. FRAGESTELLUNG.....	7
3. THEORETISCHE HINTERGRÜNDE	8
3.1. META-ANALYSE.....	8
3.1.1. <i>Einführung</i>	8
3.1.2. <i>Systematische Reviews</i>	9
3.1.3. <i>Effektgrösse und Präzision</i>	10
3.1.4. <i>„Fixed-effect“ versus „Random-effects“ Modell</i>	12
4. HAUPTTEIL	14
4.1. METHODE.....	14
4.1.1. <i>Zulassungskriterien</i>	14
4.1.2. <i>Literatursuche</i>	14
4.1.3. <i>Selektion</i>	15
4.1.4. <i>Qualitätsbeurteilung</i>	15
4.1.5. <i>Datensynthese und Analyse</i>	15
4.2. RESULTATE.....	22
4.2.1. <i>Studienauswahl</i>	22
4.2.2. <i>Studien im Detail</i>	22
4.2.3. <i>Methodologische Qualität</i>	28
4.2.4. <i>Stabilitätstraining versus manuelle Therapie</i>	28
5. DISKUSSION	31
5.1. INTERPRETATION DER RESULTATE.....	31
5.2. VERGLEICH MIT DEM REVIEW VON MACEDO ET AL. (2009).....	31
5.3. POTENTIELLER BIAS IM REVIEW-PROZESS.....	33
5.3.1. <i>“Publication Bias“</i>	33
5.3.2. <i>“Measurement Bias“ (Datensammlung)</i>	33
5.4. HETEROGENITÄT DER BEHANDLUNGSSCHEMAS.....	33
5.4.1. <i>Stabilisationstraining</i>	33
5.4.2. <i>Gruppentherapie versus Einzeltherapie</i>	34

5.5. "PRE-TREATMENT DIFFERENCES"	34
5.6. „RANDOM-EFFECTS“ MODELL BEI EINER KLEINEN ANZAHL STUDIEN.....	34
5.7. WICHTIGKEIT VON META-ANALYSEN.....	36
6. SCHLUSSFOLGERUNG	37
7. LITERATURVERZEICHNIS	38
8. DANKSAGUNG	41
9. EIGENSTÄNDIGKEITSERKLÄRUNG.....	42

1. Abstract

Grundlage

Bisher durchgeführte systematische Reviews konnten eine Wirksamkeit von Stabilisationstraining bei lumbalen Rückenschmerzen nicht eindeutig nachweisen.

Zielsetzung

Ziel dieser Arbeit war es, die Wirksamkeit von Stabilisationstraining im Vergleich zu manueller Therapie in der Behandlung von lumbalen Rückenschmerzen in einem systematischen Review von kontrollierten, randomisierten Studien zu untersuchen. Die untersuchten Variablen waren dabei Schmerz und Behinderung.

Methode

Nach relevanten Studien wurden in elektronischen Datenbanken bis zum Juli 2010 gesucht. Die Resultate bezüglich Schmerz und Behinderung wurden aus den Studien extrahiert und wenn nötig standardisiert. Wenn möglich wurden diese Resultate in einer Meta-Analyse gewichtet und zusammengefasst (pooling).

Resultate

Vier der gefundenen Studien erfüllten die Einschlusskriterien. Drei konnten für die Meta-Analyse verwendet werden. Anhand der gewichteten Zusammenfassung konnten keine signifikanten Unterschiede zwischen Stabilitätstraining und manueller Therapie bezüglich Schmerz oder Behinderung festgestellt werden. Die gewichteten Mittelwertunterschiede bezüglich Schmerzen (Visual Analog Scale, 0 - 100) betragen kurzfristig -3.6 Punkte mit einem 95% Konfidenzintervall (KI) von -9.56 bis 2.35, mittelfristig -5.58 Punkte mit einem KI von -13.97 bis 2.81 und langfristig -1.98 Punkte mit einem KI von -7.54 bis 3.58. Der Effekt der Variable Behinderung wurde standardisiert. Es ergab sich kurzfristig ein standardisierter, gewichteter Mittelwert von -0.01 Standardabweichungen mit einem KI von -0.23 bis 0.21, mittelfristig ein Mittelwert von -0.10 Standardabweichungen mit einem KI von -0.33 bis 0.13 und langfristig ein Mittelwert von -0.14 Standardabweichungen mit einem KI von -0.35 bis 0.06.

Schlussfolgerung

Was die Parameter Schmerz und Behinderung betrifft, konnte anhand der untersuchten Studien kein statistisch signifikanter Unterschied der Wirksamkeit bei

der Behandlung von lumbalen Rückenschmerzen zwischen Stabilitätstraining und manueller Therapie nachgewiesen werden.

2. Einleitung

2.1. Einführung in die Thematik

Rückenschmerzen sind ein grosses Gesundheits- und sozio-ökonomisches Problem in unserer Gesellschaft. Beschwerden des Rückens gehören zu den „häufigsten Schmerzproblemen der Bevölkerung“ (Göbel, 2001, S. 94) und bis zu 60% der Bevölkerung sind einmal im Leben davon betroffen (Göbel, 2001). Allein die volkswirtschaftlichen Kosten in der Schweiz aufgrund von Absenzen vom Arbeitsplatz gehen in die Milliarden (Läubli & Müller, 2009).

In den späten 90er Jahren wurde festgestellt, dass bei Patienten mit lumbalen Rückenschmerzen Einschränkungen der motorischen Kontrolle der tiefen Rumpfmuskulatur vorlagen. So wurde beispielsweise ein verspätetes Einsetzen des M. transversus abdominis bei Armbewegungen entdeckt. (Hodges & Richardson, 1996). Daher schien es aus theoretischer Sicht sinnvoll, eine Intervention zu erarbeiten, welche zum Ziel hatte, die Veränderungen der tiefen Rumpfmuskulatur und deren motorische Kontrolle rückgängig zu machen. Ausgehend von diesen Erkenntnissen entstanden neue Behandlungsmethoden, welche allgemein als „Motor Control Exercise“ oder Stabilisationstraining bezeichnet wurden. Dabei findet eine Progression von statischen zu dynamischen und schliesslich immer komplexeren, funktionellen Übungen statt (Richardson, Jull, Hodges, & Hides, 1999).

Es wurden einige Studien publiziert, welche unter Laborbedingungen die Grundlagen des Stabilisationstrainings zu bestätigen schienen (Hodges & Richardson, 1996, 1998). Allerdings werden heute auch die theoretischen Überlegungen, welche dem Stabilisationstraining zugrunde liegen, kontrovers diskutiert (Lederman, 2009).

Dessenungeachtet wird das Stabilisationstraining in der Praxis häufig angewendet (Stuge et al., 2004; zit. nach Ferreira P. H., Ferreira, Maher, Herbert, & Refshauge, 2006). Die Erforschung der klinischen Wirksamkeit stand im Zentrum vieler Studien. So wurden bis heute zwei systematische Reviews durchgeführt, welche die Ergebnisse der untersuchten Studien in einer Meta-Analyse gewichtet und zusammengefasst haben (Ferreira et al., 2006), (Macedo, Maher, Latimer, &

McAuley, 2009). Die klinische Wirksamkeit konnte jedoch nicht eindeutig geklärt werden.

2.2. Zielsetzung

Ziel dieser Arbeit war es, die Wirksamkeit von Stabilisationstraining im Vergleich zu manueller Therapie in der Behandlung von lumbalen Rückenschmerzen in einem systematischen Review von kontrollierten, randomisierten Studien zu untersuchen. Die untersuchten Variablen waren dabei Schmerz und Behinderung.

2.3. Fragestellung

Wie wirksam ist ein spezifisches Stabilisationstraining bezüglich Schmerz und Behinderung bei der Behandlung von lumbalen Rückenschmerzen im Vergleich zu manueller Therapie?

3. Theoretische Hintergründe

Die nachfolgenden Ausführungen beruhen auf dem Buch von Borenstein, Hedges, Higgins und Rothstein (2009).

3.1. Meta-Analyse

3.1.1. Einführung

Heutzutage wird weltweit geforscht und nahezu alle Interventionen werden in mehreren Studien auf ihre Wirksamkeit hin untersucht. Es erscheint nur schlüssig, dass man verlässlichere Aussagen zu dem wirklichen Effekt einer Intervention erhält, wenn die Evidenz zu einem Thema in ihrer Gesamtheit betrachtet wird. Die Meta-Analyse ist eine Methode zur Synthese der Resultate mehrerer, randomisierter und kontrollierter Studien zum identischen Thema.

Vor den 90er Jahren wurden Reviews vorwiegend in narrativer Form verfasst (Borenstein et al., 2009). Dabei haben die Autoren die Studien betreffend einer Fragestellung untersucht und zusammengefasst. Anschliessend gaben sie die Schlussfolgerungen wieder. Diese Form der Reviews hat einige Nachteile, welche Borenstein et al. (2009) in der Einleitung zu ihrem Buch „Introduction to Meta-Analysis“ aufzählen. Ein Nachteil ist die dieser Form zwangsläufig innewohnende Subjektivität. Zudem sind die Methoden sehr intransparent. So sind die Einschlusskriterien für die zusammengefassten Studien oft nicht eindeutig definiert. Ferner besteht die Gefahr, dass verschiedene Autoren unterschiedliche Gewichtungen vornehmen. Beispielsweise gewichtet Autor A Studien mit einer grossen Anzahl an Probanden höher, während Autor B die methodologische Qualität einer Studie als wichtiger bewertet. Ausserdem sind die Ansprüche an den Evidenzumfang, welcher ein sinnvolles Urteil zulässt, je nach Autor subjektiv. Im Allgemeinen werden die Entscheidungsfindungsprozesse, mit welchen narrative Reviewer Daten synthetisieren, selten eindeutig festgehalten, beziehungsweise die Reviewer sind sich deren nicht vollkommen bewusst. Ein weiterer Nachteil der narrativen Reviewform ist, dass diese mit zunehmendem Datenumfang und Anzahl Studien schwieriger anwendbar wird, da die Übersicht verloren geht.

3.1.2. Systematische Reviews

In Anbetracht der erwähnten Schwächen der narrativen Reviews wurden die systematischen Reviews entwickelt. Systematische Reviews bestehen grundsätzlich aus zwei Teilen: Der Literatursuche und der statistischen Synthese der Daten oder der Meta-Analyse.

Die Literatursuche und die Auswahl der Studien erfolgen aufgrund von eindeutig definierten Regeln und Kriterien. Die Definition der Einschlusskriterien birgt dabei immer noch einen gewissen Anteil an Subjektivität, allerdings werden diese transparent festgehalten und systematisch angewendet (Furlan, Pennick, Bombardier, & van Tulder, 2009).

Die Meta-Analyse stellt das Herzstück der systematischen Reviews dar. Im Gegensatz zu narrativen Reviews, bei denen die Autoren die Studien implizit nach ihrer Wichtigkeit einteilen, werden diese in einer Meta-Analyse nach im voraus definierten mathematischen Regeln gewichtet und miteinander verrechnet. Es ist so ein objektives, transparentes und reproduzierbares Resultat erreichbar. Obwohl die Autoren und Leser sich über die Interpretation der Resultate und insbesondere deren klinische Relevanz uneinig sein können, so bildet die Meta-Analyse doch eine objektive Grundlage für die Diskussion.

Systematische Reviews und Meta-Analysen werden demnach hauptsächlich mit dem Ziel durchgeführt, die vorhandene Evidenz zu einer gegebenen Fragestellung systematisch zu synthetisieren, um verlässlichere Antworten zu erhalten. Obwohl das sicher der häufigste Grund ist, gibt es noch weitere Verwendungszwecke für systematische Reviews im Forschungsprozess. So können sie eine wichtige Rolle in der Planung und im Design von neuen Forschungsprojekten spielen. Weiter kann durch systematische Reviews der Forschungsbedarf zu einer Fragestellung beurteilt werden. Zudem kann das Design verbessert werden, indem Resultate mit Instrumenten gemessen werden, die sich in früheren Studien als besonders sensibel bezüglich des untersuchten Effekts erwiesen haben.

3.1.3. Effektgrösse und Präzision

Die Grundlage für die Erstellung der Meta-Analyse bildet die Effektgrösse. Die Effektgrösse stellt das Ausmass eines Behandlungseffektes oder allgemeiner die Stärke der Beziehung zwischen zwei Variablen dar. Die Effektgrösse wird für jede Studie, die in die Meta-Analyse Eingang findet, berechnet. Anschliessend wird sie dazu verwendet, einen Überblick über die Konsistenz des Effektes über alle Studien hinweg zu erlangen, sowie einen Gesamteffekt zu errechnen. Die Effektgrösse könnte die Auswirkung einer Intervention beschreiben, so zum Beispiel die Auswirkung einer medizinischen Behandlung auf das Risiko einer Infektion oder das Ausmass einer Unterrichtsmethode auf die Prüfungsergebnisse.

Impact of Statin Dose On Death and Myocardial Infarction

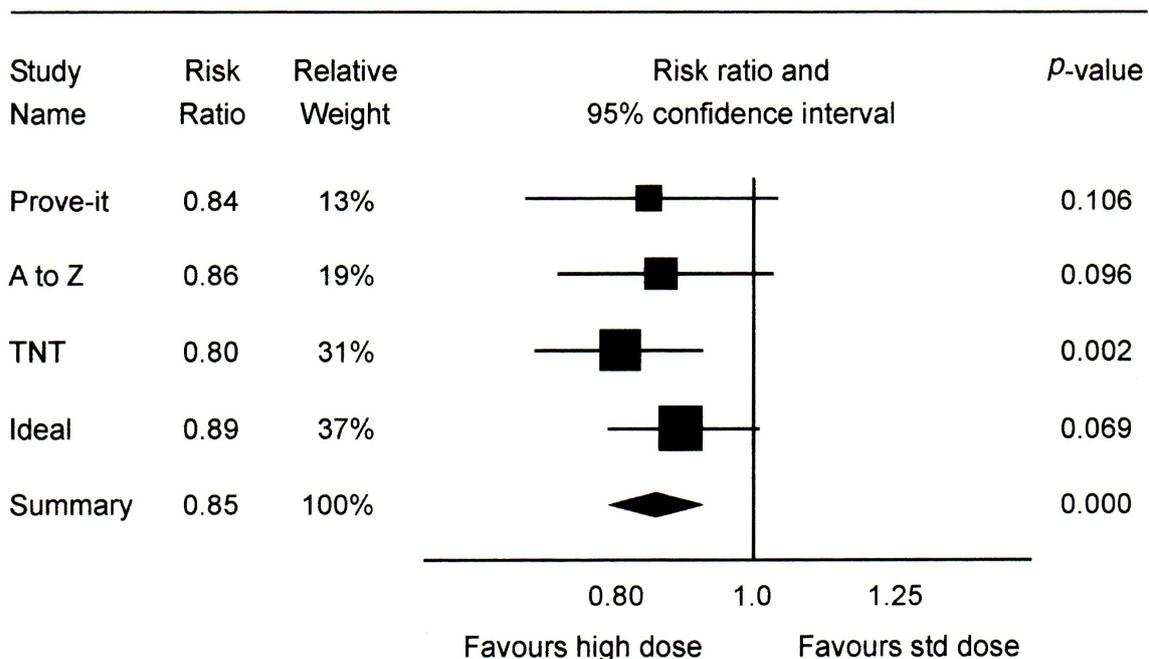


Abbildung 1: Hohe Dosis versus Standarddosis von Statin (adaptiert von Cannon et al., 2006, zit. nach Borenstein et al., 2009, S. 4)

Im Beispiel von Borenstein et al. (2009, Abbildung 1) wird die Effektgrösse anhand des Risk Ratio dargestellt. Ein Risk Ratio von 1.0 bedeutet, dass das Risiko zu sterben oder eines Herzinfarktes in beiden Gruppen gleich gross war, während ein Risk Ratio unter 1.0 ein geringeres Risiko für die Gruppe mit der hohen Dosis

bedeuten würde und umgekehrt. Die Effektgrösse jeder Studie wird durch ein Quadrat dargestellt. Das Ausmass des Effektes wird durch den Abstand von der Null-Linie wiedergegeben. Wichtig ist zu beachten, dass bei Effektgrössen, welche auf einem Verhältnis beruhen (wie im obigen Beispiel die Risk Ratio), der Effekt zwischen beiden Gruppen bei 1.0 ausgeglichen ist. Bei Effektmessungen, die auf einer Differenz beruhen (wie zum Beispiel bei der Mittelwertdifferenz) repräsentiert ein Wert von 0.0 zwei Gruppen ohne Unterschied.

Visuell wird die Effektgrösse eingebunden in ein Konfidenzintervall, das die Präzision, mit welcher die Effektgrösse in der jeweiligen Studie geschätzt wurde, wiedergibt. Allgemein ist die Präzision von Studien primär von der Anzahl Teilnehmer abhängig, wobei eine grössere Anzahl präzisere Schätzungen liefert. Weitere Faktoren, welche die Präzision beeinflussen können, liegen im Studiendesign. Studien, die sogenannte „matched groups“ oder gepaarte Gruppen verwenden, sind in der Regel präziser als randomisierte Gruppen. Dies gilt jedoch nur wenn die Paarungen sinnvoll sind. Das bedeutet, dass die Paare bezüglich der Faktoren, welche den Effekt beeinflussen könnten, ähnlich sind. (In kontrollierten Studien mit „matched groups“ werden alle Probanden vor der Zuteilung zu Gruppen, in möglichst ähnliche Paare, betreffend festgelegter Faktoren, eingeteilt. Zum Beispiel Alter, Geschlecht, Gewicht usw.). Je ein Proband pro Paar wird anschliessend einer Gruppe zugeteilt.

Eine weitere Information, welche die Abbildung 1 beinhaltet, sind die Flächen der Quadrate, welche die Effektgrössen der Studien markieren. Je grösser die Fläche des Quadrates, desto grösseres Gewicht wurde der Studie zugeteilt. Studien mit grösserem Gewicht haben einen grösseren Anteil am mathematisch berechneten Gesamteffekt. Allgemein erhalten Studien mit einer grösseren Präzision auch ein grösseres Gewicht. Wie schon besprochen, ist die Präzision vor allem von der Probandenanzahl abhängig. Das Gewicht wird jedoch nicht nur durch die Präzision bestimmt.

In einer Meta-Analyse müssen auch Annahmen über die in den Studien untersuchten Effekte getroffen werden. Es kann zum einen davon ausgegangen werden, dass allen Studien nur ein wirklicher Effekt zugrunde liegt oder zum anderen jede Studie einen anderen, jedoch ähnlichen wirklichen Effekt beschreibt. Diese Annahme beeinflusst die Zuweisung einer Gewichtung. Darauf wird im nächsten Kapitel (3.1.4) genauer eingegangen.

3.1.4. „Fixed-effect“ versus „Random-effects“ Modell

Meta-Analysen basieren hauptsächlich auf dem „Fixed-effect“ Modell oder dem „Random-effects“ Modell.

Beim „Fixed-effect“ Modell wird die Annahme getroffen, dass es genau *einen* Effekt gibt, der allen Studien des Reviews zugrunde liegt. Alle Unterschiede in den Resultaten der Studien sind folglich durch Stichprobenfehler zu erklären. Da es sich um nur einen Effekt handelt, wird bewusst der Singular, „Fixed-effect“ verwendet.

Im Gegensatz dazu erlaubt das „Random-effects“ Modell, dass die Effekte von Studie zu Studie variieren. Es wird also angenommen, dass es nicht nur einen Effekt gibt, der den untersuchten Studien zugrunde liegt. So ist vorstellbar, dass der Effekt zum Beispiel je nach Probandenzusammensetzung mit unterschiedlichem Alter, Geschlecht oder Gesundheitszustand auch unterschiedlich ausfallen wird. Weiter könnte auch ein grösserer (oder kleinerer) Effekt bei Studien mit intensiveren Interventionen vorliegen. Wenn es möglich wäre, eine unendliche Anzahl von Studien (basierend auf den Einschlusskriterien der Analyse) durchzuführen, würden sich die Effektgrößen um einen Mittelwert streuen. Die Effektgrößen der Studien, die wirklich durchgeführt wurden, stellen eine zufällige Stichprobe dieser Verteilung dar. Daher wird dieses Modell „Random-effects“ (random=zufällig) genannt. Ebenfalls ist zu beachten, dass nun der Plural („effects“) verwendet wird, da von einer Vielzahl von Effekten ausgegangen wird.

Wie im vorhergehenden Kapitel beschrieben (3.1.3), werden die Effektgrößen der untersuchten Studien in einer Meta-Analyse aufgrund der Probandengröße gewichtet und anschliessend eine Schätzung des Gesamteffektes („summary effect“)

berechnet. Bei der Gewichtung der Studien gibt es einen wichtigen Unterschied zwischen den beiden Modellen zu beachten. Wenn das „Fixed-effect“ Modell angewendet wird, erhalten die Resultate von kleineren Studien bezüglich der Probandenanzahl viel weniger Gewicht, da in den grösseren Studien viel bessere Informationen über den *selben* Effekt vorliegen.

Im Gegensatz dazu ist es im „Random-effects“ Modell nicht das Ziel, einen wirklichen Effekt zu schätzen, sondern den Mittelwert einer Streuung von Effekten. Da alle Studien einen anderen Effekt abschätzen, muss sichergestellt werden, dass sämtliche Effekte auch in die Zusammenfassung einfliessen. Folglich können die Ergebnisse von kleinen Studien nicht so drastisch abgewertet werden, wie das bei Anwendung des „Fixed-effects“ Modells vollzogen würde. Umgekehrt können grössere Studien nicht so stark gewichtet werden. Die Schätzung des Effektes dieser Studie mag zwar präziser sein, aber sie liefert nur Informationen über einen bestimmten Effekt, der von allen anderen Studien nicht untersucht wurde. Das Ziel ist es, den mittleren Effekt einer Auswahl von Studien abzuschätzen. Dabei gilt es zu verhindern, dass das Endergebnis nicht übermässig durch eine einzelne Studie beeinflusst wird.

4. Hauptteil

4.1. Methode

4.1.1. Zulassungskriterien

Für diese Arbeit ausgewählte Studien mussten bestimmte Bedingungen erfüllen. So musste es sich um randomisierte, kontrollierte Studien handeln. Weiter konnten nur Studien, die in englischer, deutscher, französischer oder italienischer Sprache verfasst worden sind, in Betracht gezogen werden, da nur diese Sprachen vom Autor verstanden werden. Weiter musste es sich um erwachsene Teilnehmer handeln, die Symptome im lumbalen Rücken aufwiesen. Zwingend war weiter, dass mindestens eine Gruppe ein spezifisches Stabilisationstraining vollzog. Die Studien waren auch gültig, wenn das Stabilisationstraining so nicht erwähnt wurde, jedoch sinnvoll beschrieben wurde (Training der entsprechenden Muskulatur, also der Mm. multifidus und transversus). Mindestens eine weitere Kontrollgruppe musste mittels manueller Therapie behandelt werden, da dies der Fragestellung entsprach. Das Stabilisationstraining durfte in der Interventionsgruppe nicht nur als Zusatz angewendet werden. Dies war aus Abgrenzungsgründen zu anderen Arbeiten nötig. Der Interventionseffekt musste mindestens bezüglich den Parametern Schmerz und Behinderung überprüft worden sein.

4.1.2. Literatursuche

Die Suche nach relevanten Artikeln fand rechnergestützt statt. Für die Literatursuche wurden die Datenbanken Medline, PEDro und CINAHL verwendet. Die Schlagwörter bezüglich randomisierten kontrollierten Studien und lumbalen Rückenschmerzen (Low Back Pain) richteten sich nach den Empfehlungen der Cochrane Back Review Group (Furlan et al., 2009). Die Begriffe für die motorische Kontrolle orientierten sich an den Vorgaben von Ferreira et al. (2009). Diese wurden soweit eingeschränkt, um nur Studien in Bezug auf lumbale Rückenschmerzen zu erhalten. Dabei wurden die Begriffe in der Suche mit OVID mittels Booleschen Operatoren wie folgt sinnvoll verknüpft: (specific OR stabili\$ OR segment\$ OR multifidus OR transversus).mp AND (exercise OR train\$).mp.

Ausserdem wurden die Literaturverzeichnisse von bereits publizierten Reviews und geeigneten Studien durchsucht und mit einbezogen („citation tracking“).

4.1.3. Selektion

Die Resultate der Datenbankrecherche wurden anschliessend aufgrund der in 4.1.1 erwähnten Kriterien ausgewählt. Es wurde eine Checkliste erstellt, um sicherzustellen, dass kein Kriterium vergessen wurde.

4.1.4. Qualitätsbeurteilung

Die Qualität der Studien wurde mittels der PEDro Skala bewertet. Die Reliabilität der PEDro Skala wurde als akzeptabel beschrieben (Maher, Sherrington, Herbert, Moseley, & Elkins, 2003). Die Ergebnisse der eigenen Bewertung wurden mit denen der PEDro Datenbank und anderen Reviews verglichen (Ferreira et al., 2006), (Macedo et al., 2009). Die eigene Beurteilung war nötig, da bei einigen untersuchten Studien nur Teile verwendet wurden, was zum Beispiel aufgrund der veränderten Gruppengrössen zu anderen Resultaten in der PEDro Score führen könnte. Bei Differenzen, die sich nicht klären liessen, wurden die Ergebnisse der PEDro Datenbank verwendet, die bereits von mindestens zwei Fachpersonen beurteilt wurden.

4.1.5. Datensynthese und Analyse

Zur Analyse der Resultate der untersuchten Studien wurden die Ergebnisse bezüglich Behinderung und Schmerz extrahiert. Falls einer der untersuchten Parameter in allen untersuchten Studien mit den selben Instrumenten und der selben Skala untersucht wurde, so wurden die Rohdaten direkt zur Meta-Analyse verwendet. Wenn die Skalen jedoch unterschiedlich waren, so wurden diese zuerst standardisiert (vergleiche unten).

Die Zeitpunkte der Untersuchungen wurden aufgrund der Empfehlungen der Cochrane Back Review Group (Furlan et al., 2009) gewählt. Es werden dabei kurzfristige („measured closest to 8 weeks“), mittelfristige („measured closest to 6 months“) und langfristige („measured closest to 12 months“) Resultate unterschieden.

Die Auswahl der untersuchten Parameter Behinderung und Schmerz wurden so gewählt, dass die Ergebnisse als Mittelwerte in kontinuierlicher Form vorlagen. So konnten diese Mittelwerte, wie nachfolgend beschrieben, gewichtet und kombiniert werden. Die Ergebnisse werden sowohl numerisch als auch visuell in einem Forest-Plot dargestellt (Bei einem Forest-Plot handelt es sich um eine graphische Darstellung der Resultate aller involvierten Studien in einer systematischen Übersicht). Zudem wird das durch die Meta-Analyse errechnete Gesamtergebnis abgebildet. Die Ergebnisse der qualitativen Untersuchung mittels der PEDro Scale fließen dabei der Reihenfolge nach im Forest-Plot ein. Die Studien werden absteigend gemäss ihrer Anzahl Punkte angeordnet. Bei Punktgleichheit erfolgt die Anordnung alphabetisch absteigend, ausgehend vom Nachnamen des erstgenannten Autors.

Wenn in Studien nur der Mittelwert aber keine Standardabweichungen angegeben wurden, so wurde ein Schätzwert aufgrund der Empfehlungen der Cochrane Back Review Group (Furlan et al., 2009) berechnet. Falls in einer Studie keine Mittelwerte aufgeführt waren (sondern zum Beispiel der Median), wurden die Autoren kontaktiert.

Modell-Wahl

Wie in Kapitel 3.1.4 erwähnt, stehen für eine Meta-Analyse hauptsächlich zwei Modelle zur Erstellung der gewichteten Zusammenfassung („pooling“) zur Verfügung. In dieser Arbeit wurde das „Random-effects“ Modell angewendet. Da alle Studien von verschiedenen, unabhängig von einander arbeitenden Wissenschaftlern durchgeführt wurde, ist es unwahrscheinlich, dass die Studien funktionell äquivalent waren, wie dies von Borenstein et al. (2009) zur Anwendung des „Fixed-effects“ Modells gefordert wird. Die Interventionen waren zu heterogen, als dass von einem einzigen gemeinsamen Effekt ausgegangen werden kann. Die Wahl des „Random-effects“ Modells wird auch von den „Guidelines for Systematic Reviews in the Cochrane Back Review Group“ empfohlen (Furlan et al., 2009).

Effektgrössenberechnung

Die nachfolgenden Berechnungen beruhen auf den Ausführungen von Borenstein et al. (2009).

Differenz der Mittelwerte D (raw mean difference) bei unabhängigen Gruppen

Wenn die Resultate in einer aussagekräftigen, wohl bekannten Skala publiziert wurden und alle Studien dieselbe Skala verwendeten, konnten direkt die Rohdaten zur Meta-Analyse verwendet werden. Wenn also Mittelwerte von zwei Gruppen angegeben wurden, wobei μ_n für wirkliche mittlere Effektgrössen der Populationen steht, dann gilt für die Differenz der Mittelwerte Δ :

$$\Delta = \mu_1 - \mu_2$$

Nun kann die Differenz der Mittelwerte Δ anhand der Studienresultate abgeschätzt werden. Dabei stehen \bar{X}_n für die Mittelwerte der Stichproben. Die Schätzung von Δ aufgrund der Stichproben wird als D bezeichnet und berechnet sich wie folgt:

$$D = \bar{X}_1 - \bar{X}_2$$

Bezeichne nun S_n die Standardabweichung der Stichproben und n_1 und n_2 die Stichprobengrössen, dann errechnet sich die Varianz von D wie folgt:

$$V_D = \frac{n_1 + n_2}{n_1 n_2} S_{pooled}^2 \quad (\text{wenn } \sigma_1 = \sigma_2 = \sigma)$$

wobei gilt:

$$S_{pooled} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

S_{pooled} beschreibt dabei die kombinierte Standardabweichung innerhalb der Gruppen. Es wird dabei davon ausgegangen, dass die Standardabweichungen der Populationen gleich sind, was bei einer guten Homogenität der beiden Gruppen („baseline comparability“) der Fall sein sollte.

Der Standardfehler von D ergibt sich somit als:

$$SE_D = \sqrt{V_D}$$

Standardisierte Mittelwertunterschiede d (standardized mean difference)

Wenn die Studienresultate in unterschiedlichen oder weniger bekannten Skalen vorlagen oder die Studien verschiedene Messinstrumente für dieselbe untersuchte Variable verwendeten, mussten diese vor der Meta-Analyse erst standardisiert werden. Dabei wird die Differenz der Mittelwerte in jeder Studie durch die Standardabweichung der Studie geteilt. Daraus ergibt sich ein Wert, der über alle Studien hinweg verglichen werden kann. Dieser Ansatz wurde erstmals von Cohen (1969, zit. nach Borenstein et al., 2009) vorgeschlagen.

Nehmen wir an, eine Studie vergleiche zwei unabhängige Gruppen. Weiter sei μ_n der wirkliche Mittelwert der Population n und σ_n die wirkliche Standardabweichung der Population n , dann ergibt sich die standardisierte mittlere Differenz δ aus:

$$\delta = \frac{\mu_1 - \mu_2}{\sigma} \text{ wenn } \sigma_1 = \sigma_2 = \sigma$$

Die standardisierte mittlere Differenz δ kann nun aufgrund von zwei Studien mit unabhängigen Gruppen geschätzt werden. Diese Schätzung von δ (dem Populationsparameter) aufgrund der Stichprobe wird als Cohens d bezeichnet:

$$d = \frac{\bar{X}_1 - \bar{X}_2}{S_{pooled}}$$

Dabei gilt für kombinierte Standardabweichung S_{pooled} wie zuvor:

$$S_{pooled} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

Die Varianz von d lässt sich bestimmen durch:

$$V_d = \frac{n_1 + n_2}{n_1 n_2} + \frac{d^2}{2(n_1 + n_2)}$$

wobei sich der Standardfehler von d ergibt aus:

$$SE_d = \sqrt{V_d}$$

Bewertung der Effektgrößen

Die Resultate der untersuchten Variable Schmerz muss aufgrund der einheitlichen und weit verbreiteten Skala (Visual Analog Scale, VAS) nicht standardisiert werden. Bei lumbalen Rückenschmerzen werden Differenzen der Mittelwerte von 30% oder mehr als klinisch relevant betrachtet (Ostelo et al., 2008).

Resultate, welche aufgrund der Heterogenität ihrer Messinstrumente zuerst standardisiert werden mussten, werden anhand der Empfehlungen von Cohen nach der Grösse ihres Effektes bewertet (Cohen, 1988). Er bezeichnet dabei einen Effekt von $d=0.2-0.3$ als klein, einen Effekt von $d=0.5$ als mittel und einen Effekt von $d=0.8$ als gross.

Gewichtung mittels dem „Random-effects“ Modell

Wie im Theorieteil beschrieben, wird beim „Random-effects“ Modell davon ausgegangen, dass mehrere wirkliche Effekte vorliegen. Der wirkliche Effekt einer Studie sei θ . Diese Effekte haben einen Mittelwert μ und eine Verteilung von der Standardabweichung ζ . Aufgrund des Stichprobenfehlers ε ist der beobachtete Effekt Y aber nicht unbedingt gleich θ . Der beobachtete Effekt Y_i jeder beteiligten Studie setzt sich also zusammen aus dem Mittelwert aller Studien μ , der Abweichung der Studie i von diesem Mittelwert und der Abweichung der Studie von ihrem wirklichen Effekt. In einer Formel dargestellt gilt also:

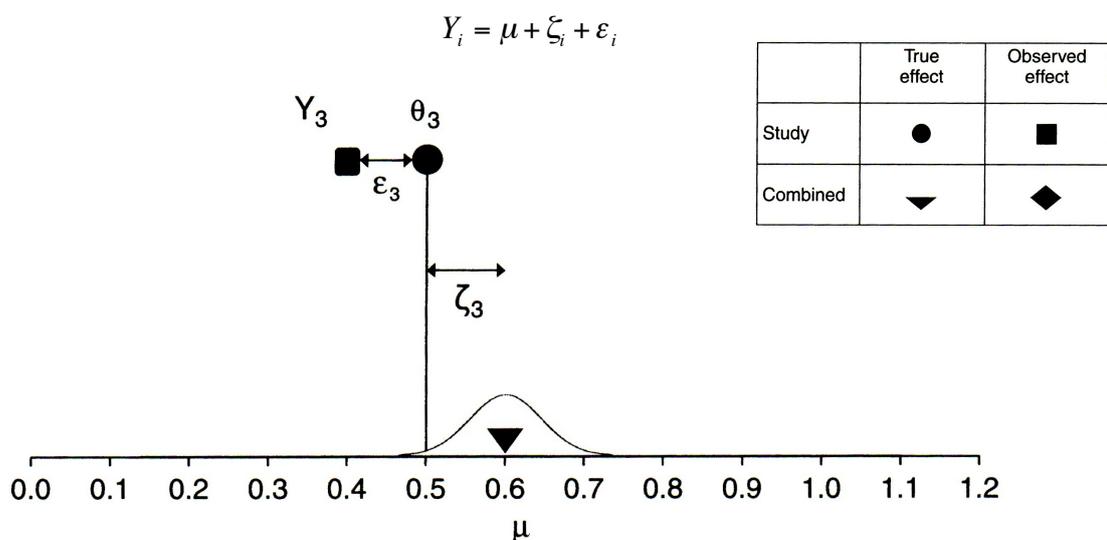


Abbildung 2: „Random-effects“ Modell – wirkliche und beobachtete Effekte einer Studie (Borenstein et al., 2009, S. 71)

Die Distanz von μ bis θ ist abhängig von der Standardabweichung der Verteilung der wirklichen Effekte aller Studien, welche als τ bezeichnet wird (oder τ^2 für die Varianz). Dieses τ wird geschätzt mittels T :

$$T^2 = \frac{Q - df}{C}$$

wobei gilt:

$$Q = \sum_{i=1}^k W_i Y_i^2 - \frac{\left(\sum_{i=1}^k W_i Y_i \right)^2}{\sum_{i=1}^k W_i}$$

$$df = k - 1 \quad (k = \text{Anzahl Studien})$$

$$C = \sum W_i - \frac{\sum W_i^2}{\sum W_i}$$

Als W_i wird dabei die Gewichtung einer Studie i , wie sie im „Fixed-effect“ Modell berechnet wird, bezeichnet.

$$W_i = \frac{1}{V_{Y_i}}$$

Wobei V_{Y_i} der („within-study“) Varianz der Studie i entspricht.

Das Gewicht jeder Studie im „Random-effects“ Modell wird jetzt noch durch die Schätzung der „between-study“ Varianz ergänzt (zur Unterscheidung vom „Fixed-effect“ Modell wird ein * angefügt):

$$W_i^* = \frac{1}{V_{Y_i} + T^2}$$

Der gewichtete Mittelwert M^* entspricht somit:

$$M^* = \frac{\sum_{i=1}^k W_i^* Y_i}{\sum_{i=1}^k W_i^*}$$

Mit der Varianz:

$$V_{M^*} = \frac{1}{\sum_{i=1}^k W_i^*}$$

Die Resultate wurden grafisch in Forest-Plots dargestellt. Zu deren Erstellung wurde das Computerprogramm Review Manager (Version 5.0.24, The Cochrane Collaboration, 2008) verwendet.

Δ	mittlere Differenz der Population
μ_n	wirklicher Populationsmittelwert der Gruppe n
\bar{X}_n	Stichprobenmittelwert der Gruppe n
D	Differenz der Stichprobenmittelwerte
d	standardisierte Differenz der Stichprobenmittelwerte (Cohens d)
S_n	Standardabweichung der Stichprobe der Gruppe n
S_{pooled}	kombinierte Standardabweichung
n	Stichprobenumfang
V	Varianz
SE	Standardfehler
δ	standardisierte, wirkliche mittlere Differenz
σ_n	wirkliche Standardabweichung der Population n
θ_i	wirklicher Effekt der Studie i
Y_i	beobachteter Effekt der Studie i
ζ_i	Abweichung des wirklichen Effektes der Studie i (θ_i) vom wirklichen Mittelwert der Population μ
ε_i	Stichprobenfehler der Studie i
τ	Standardabweichung der Verteilung der wirklichen Effekte („between-study“)
T	Schätzung von τ
W_i	Gewichtung der Studie i
M	gewichteter Mittelwert

4.2. Resultate

4.2.1. Studienauswahl

Die anfängliche elektronische Literatursuche ergab 132 Treffer. Davon kamen 15 Studien als potentiell zulässig in Frage. Diese wurden genau auf die Zulassungskriterien überprüft, wodurch sich schliesslich vier Studien (Critchley, Ratcliffe, Noonan, Jones, & Hurley, 2007), (Ferreira et al., 2007), (Goldby, Moore, Doust, & Trew, 2006), (Rasmussen-Barr, Nilsson-Wikmar, & Arvidsson, 2003) als zulässig für die Verwendung in dieser Arbeit erwiesen.

4.2.2. Studien im Detail

In den folgenden Ausführungen wird der Fokus auf die für diese Arbeit relevanten Interventionen (Stabilisationstraining und manuelle Therapie) und Messgrössen (Schmerz und Behinderung) gelegt.

Rasmussen-Barr et al. (2003)

Patienten

Die Autoren untersuchten 47 Patienten mit sub-akuten, chronischen oder rezidiven lumbalen Rückenschmerzen, die sich in einer Klinik in Stockholm behandeln liessen und sich zur Teilnahme an der Studie bereit erklärt haben. Die Patienten mussten zwischen 18-60 Jahre alt sein. Die Rückenschmerzen mussten mindestens seit sechs Wochen bestehen und konnten mit oder ohne Ausstrahlungen auftreten. Die wichtigsten Ausschlusskriterien waren: Früheres Stabilisationstraining oder Operationen an der Wirbelsäule, manuelle Therapie in den vorangegangenen drei Monaten und bekannte Diskushernien.

Interventionen

Die Patienten wurden zufällig in zwei Gruppen eingeteilt. Diese Gruppen wurden jeweils mit manueller Therapie oder einem spezifischen Stabilitätstraining behandelt. Die Stabilitätstrainingsgruppe erhielt während sechs Wochen ein wöchentliches Training von 45 Minuten in individuellen Sitzungen. Die Therapie wurde mittels Palpation und aufblasbaren Druckkissen überwacht. Auf die Anwendung eines Ultraschallgerätes wurde verzichtet. Die Patienten wurden dazu aufgefordert, die

erlernten Übungen täglich für 15 Minuten auszuführen. Die Einhaltung des Übungsprogrammes wurde mittels eines Tagebuches kontrolliert.

Die manuelle Therapiegruppe wurde ebenfalls in wöchentlichen Einzelsitzungen für 45 Minuten und während sechs Wochen behandelt. Die Therapie umfasste eine Kombination aus Muskeldehnung, segmentaler Traktion und Weichteilmobilisation. Die Patienten wurden zudem dazu aufgefordert, ihre gewohnten Aktivitäten und Übungen fortzuführen.

Resultate

Die Analyse der Differenz innerhalb jeder Gruppe nach der Behandlungsperiode zeigte eine signifikante Verbesserung der Stabilitätsgruppe bezüglich Schmerz (VAS) und Behinderung (Oswestry Disability Index, ODI, 0–100, 0=best). Innerhalb der Manipulationsgruppe konnte langfristig keine signifikante Verbesserung bezüglich Schmerz oder Behinderung festgestellt werden. Der Vergleich zwischen den Gruppen („between-groups analysis“) zeigte keine signifikanten Unterschiede bezüglich Schmerz oder Behinderung. Weiter wurde festgestellt, dass sich nach 3 und 12 Monaten signifikant mehr Patienten der manuellen Therapiegruppe wieder in Behandlung befanden ($P < 0.05$, jeweils 50% vs. 11%).

Kritische Diskussion

Ein Kriterium, welches in der PEDro Scale begutachtet wird ist die „Drop-out“ beziehungsweise „Follow-up“ Rate. Diese muss unter 15%, der ursprünglich in Gruppen eingeteilten Anzahl Teilnehmer sein. Nun muss diese Zahl nur zu einem der Zeitpunkte der Nachbeurteilung erzielt worden sein. Diese Raten liegen in der Studie von Rasmussen-Barr et al. (2003) nur bei der Messung direkt nach der Behandlung unter den von der PEDro Scale geforderten 15%. Im weiteren Verlauf konnte jedoch kein adäquater „Follow-up“ erzielt. Ein „Loss-to-follow-up“ von 21,4% mittelfristig und 23,8% die langfristigen Resultate betreffend, schwächt die Aussagekraft der Studie erheblich.

Ferreira et al. (2006)

Patienten

Ferreira et al. untersuchten in ihrer Studie insgesamt 240 Patienten, die sich in drei Spitälern in Sydney aufgrund lumbaler Rückenschmerzen behandeln liessen. 160 Probanden wurden dabei in die für diese Arbeit relevanten Gruppen eingeteilt. Die Patienten litten seit mindestens drei Monaten unter unspezifischen lumbalen Rückenschmerzen. Ausschlusskriterien waren unter anderem frühere Operationen an der Wirbelsäule. Patienten die berichteten, unter Osteoarthritis oder einer Diskusläsion (Prolaps, Protrusion oder Herniation) ohne neurologische Kompression zu leiden, waren ausdrücklich zugelassen.

Intervention

Die Therapien fanden während einer Periode von acht Wochen statt. In der Stabilisationsgruppe wurden immer 12 Sitzungen abgehalten. In der Manipulationsgruppe waren es maximal 12. Der Therapeut konnte die Behandlung selbstständig beenden. Beide Gruppen erhielten Einzeltherapien. Im Stabilitätstraining wurde die spezifische Rumpfmuskulatur trainiert. Das Training wurde mittels Ultraschallmessung überwacht. Die spinale Manipulation umfasste Gelenkmobilisation und Manipulation der Wirbelsäule oder des Beckens.

Resultate

Die Studienergebnisse bezüglich Schmerz (VAS) und Behinderung (Roland Morris Disability Questionnaire, 0-24, 0=best) wurden nach 8 Wochen, 6 und 12 Monaten gemessen. Sowohl die Schmerzen und die Behinderung reduzierten sich in beiden Gruppen schon nach 8 Wochen signifikant. Die Ergebnisse zeigten in der Stabilitätsgruppe tendenziell leicht bessere Resultate als in der manuellen Therapiegruppe. Die Unterschiede in den Ergebnissen waren jedoch zu keinem der gemessenen Zeitpunkte statistisch signifikant („between-group differences“).

Kritische Diskussion

Die Qualität der Studie von Ferreira et al. war grundsätzlich gut, was auch in einer guten Bewertung in der PEDro Score bestätigt wurde (siehe unten). Anzumerken ist jedoch, dass die manuelle Therapie nur während maximal 12 Sitzungen durchgeführt wurde, wobei der behandelnde Therapeut die Behandlung nach eigener

Einschätzung beenden konnte. Im Gegensatz dazu wurden in der Stabilisationsgruppe immer alle 12 Sitzungen abgehalten.

Goldby et al. (2006)

Patienten

Die Teilnehmer dieser Studien litten an chronischen, lumbalen Rückenschmerzen, wobei die aktuelle Episode mindestens seit 12 Wochen bestand. Die wichtigsten Ausschlusskriterien waren Anzeichen von neurologischen Defiziten oder vorhergehende Operationen an der Wirbelsäule. Insgesamt wurden 302 Probanden in drei Interventionsgruppen eingeteilt (Stabilisationstraining, manuelle Therapie, Rückenschule).

Intervention

Die Therapiensitzungen dauerten eine Stunde und fanden wöchentlich während zehn Wochen statt. Das Stabilisationstraining wurde in Gruppen von maximal 12 Personen abgehalten. Die manuelle Therapie wurde in Form von Einzelsitzungen durchgeführt. Die Therapeuten der manuellen Therapiegruppen konnten die Behandlung auch schon früher als erst nach zehn Wochen abbrechen, wenn sie diese für nicht mehr nötig betrachteten.

Resultate

Die Messung der Therapieerfolge wurde mittels den Faktoren Schmerz (VAS) und Behinderung (ODI) durchgeführt. Die Datenerhebung fand vor der ersten Intervention sowie nach 3, 6, 12 und 24 Monaten statt. Von den 302 Probanden führten 75 (24.8%) das Interventionsprogramm nicht zu Ende. Sie werden als „Drop-outs“ bezeichnet. Die Verteilung der „Drop-outs“ bezüglich der Interventionsgruppen war nicht aus der Studie ersichtlich. Die Resultate von 14 (4.6%) weiteren Probanden, welche das Interventionstraining vollständig durchgeführt hatten, konnten aufgrund von fehlender Unterlagen nicht für die Auswertung verwendet werden („Loss-to-follow-up“).

Im Vergleich der beiden Gruppen konnte die Stabilisationsgruppe betreffend der Variable Schmerz (VAS) nach sechs Monaten statistisch signifikant bessere Werte aufweisen ($P=0.009$). Nach 12 Monaten lies sich dieser Unterschied jedoch nicht

mehr nachweisen ($P=0.7$). Bezüglich dem untersuchten Parameter Behinderung (ODI) konnte zu keinem der gemessenen Zeitpunkte ein signifikanter Unterschied zwischen den Gruppen nachgewiesen werden.

Kritische Diskussion

Die Stratifizierung der Patienten vor der Einteilung in Gruppen ist sehr gut und wird gut dokumentiert und begründet. Die „baseline-comparability“ ist bezüglich vielen wichtigen Parametern, wie dem Geschlecht oder der Altersstruktur, gut gewährleistet und überprüfbar. Allerdings sind die Gruppen, was das Ausmass der Schwere der Symptome wie Schmerz betrifft, deutlich unterschiedlich. Aufgrund des beobachteten natürlichen Verlaufes von Schmerzen bei lumbalen Rückenschmerzen ist davon auszugehen, dass bei einem höheren Anfangsniveau von Schmerzen eine stärkere Abnahme der Schmerzen pro Zeiteinheit auftritt, was anschliessend in einem grösseren Delta und somit in einem grösseren Effekt münden würde. Ein gleicher Effekt würde nur auftreten, wenn die Abnahme der Schmerzen linear und gleich schnell stattfinden würde.

Weiter geht aus der Studiendokumentation nicht hervor, in welcher Frequenz die manuelle Therapie abgehalten wurde, wobei es beim Stabilisationstraining immer wöchentlich war. Mögliche Unterschiede zwischen den Gruppen, könnten also auch aufgrund von unterschiedlichen Intensitäten der Behandlung zurückzuführen sein. Zudem sah es das Behandlungs-Protokoll vor, dass der Behandlungsabschluss in der manuellen Therapiegruppe frei durch den behandelnden Therapeuten bestimmt werden durfte. Die Behandlungsdauer könnte sich also beträchtlich von der der Stabilisationsgruppe unterscheiden, was ebenfalls Unterschiede zwischen den Gruppen verursachen könnte.

Ein weiterer wichtiger Kritikpunkt ist die hohe „Drop-out“ beziehungsweise „Loss-to-follow-up“ Rate. Von den insgesamt 302 ursprünglich in Gruppen eingeteilten Probanden konnten nur 213 mindestens einmal bezüglich der untersuchten Variablen erfasst werden, was einem Verlust von sehr hohen 29.5% entspricht.

Critchley et al. (2007)

Patienten

Critchley et al. untersuchten insgesamt 212 Patienten, welche bereits seit mindestens 12 Wochen an Rückenschmerzen litten und von Spezialisten oder Hausärzten zur Behandlung Physiotherapie verordnet bekamen. Die Patienten waren mindestens 18 Jahre alt. Die hauptsächlichen Ausschlusskriterien waren frühere Operationen an der Wirbelsäule oder Physiotherapie während den vorangegangenen sechs Monaten.

Interventionen

Die Patienten wurden in drei Interventionsgruppen eingeteilt. Zwei davon waren für die vorliegende Arbeit von Relevanz. Die erste Gruppe (n=71) erhielt individuelle Physiotherapie in Form von Gelenkmobilisation und Gelenkmanipulation sowie Massagen. Es handelte sich dabei um maximal zwölf Sitzungen von 30 Minuten Dauer. Aktive Übungen erhielten die Patienten als Heimübungen. Die zweite Gruppe (n=72) wurde mit Stabilisationstraining behandelt. Diese fand sowohl individuell als auch in Gruppen statt. Der Behandlungsumfang betrug maximal acht Sitzungen von 90 Minuten Dauer. Die anteilmässige Aufteilung von Gruppen- und Einzeltherapie ging nicht aus dem Artikel hervor.

Resultate

Die Effekte der Intervention wurden bezüglich Behinderung mittels des RDQ und bezüglich Schmerz mittels VAS überprüft. Diese Werte wurden nach 6, 12 und 18 Monaten ermittelt. Der Vergleich zwischen den Gruppen ergab sowohl bezüglich Behinderung sowie Schmerz zu keinem Zeitpunkt statistisch signifikante Unterschiede.

Kritische Beurteilung

Die erste Bewertung der Ergebnisse fand erst nach 6 Monaten statt. Dies ist auch im Vergleich zu den Empfehlungen der Cochrane Back Review Group (Furlan et al., 2009) relativ spät. Infolgedessen fehlen Informationen über die Wirksamkeit einer Therapie, den kurz- bis mittelfristigen Genesungsvorgang zu unterstützen oder zu beschleunigen.

In der manuellen Therapiegruppe wurden auch Heimübungen zum spezifischen Rumpfraining instruiert. Es handelte sich also nicht um eine rein manuelle Therapie. Obwohl dies wahrscheinlich der Realität näher kommt, so ist trotzdem vorstellbar, dass dadurch ein möglicher Effekt des spezifischen Stabilitätstrainings weniger deutlich hervortritt. Je nach Ausführung und Art der Rumpfübungen, die in der Studie nicht näher beschrieben wurden, können diese auch eine Wirkung ähnlich dem Stabilisationstraining entfalten.

4.2.3. Methodologische Qualität

Die Ergebnisse der Beurteilung der Studien mittels der PEDro Scale sind aus der Tabelle 1 ersichtlich.

	Eligibility criteria*	Random allocation	Concealed allocation	Baseline Comparability	Blinded subjects	Blinded therapists	Blinded assessors	Adequate follow-up	Intention-to-treat analysis	Between-group comparison	Point estimates and variability	Total
Ferreira et al. (2007)	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes	8
Critchly et al. (2007)	Yes	Yes	Yes	Yes	No	No	Yes	No	Yes	Yes	Yes	7
Rasmussen-Barr et al. (2003)	Yes	Yes	No	Yes	No	No	No	Yes	No	Yes	Yes	5
Goldby et al. (2006)	Yes	Yes	No	No	No	No	Yes	No	No	Yes	Yes	4

Tabelle 1: Beurteilung mittels der PEDro Scale, *Die Zulassungskriterien zählen nicht zum PEDro Gesamt-Score.

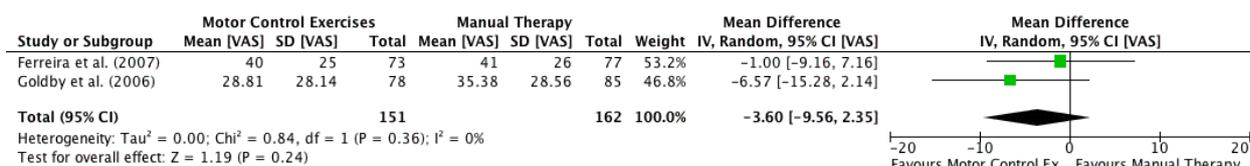
Die Resultate der methodologischen Untersuchung werden in den Forest-Plots visuell anhand der Reihenfolge wiedergegeben. Die Studien werden absteigend gemäss ihrer Anzahl Punkte angeordnet.

4.2.4. Stabilitätstraining versus manuelle Therapie

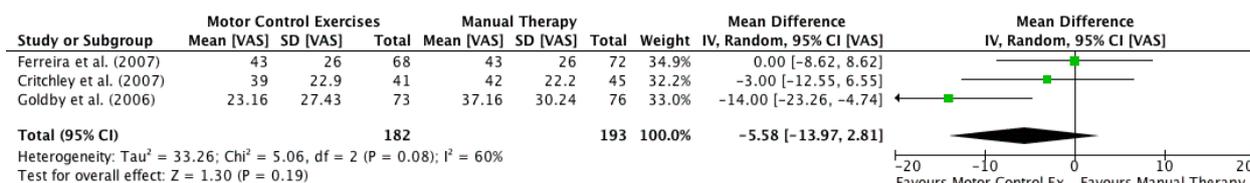
Von den vier gefundenen Studien konnten drei (Critchley et al., 2007), (Ferreira et al., 2007), (Goldby et al., 2006) für die Meta-Analyse verwendet werden. Von einer Studie (Rasmussen-Barr et al., 2003) lagen unzureichende Daten vor, wobei eine Nachfrage bei den Autoren bei Abgabe dieser Arbeit noch nicht beantwortet wurde.

Die kombinierten Resultate der Meta-Analyse zeigten bezüglich der untersuchten Parameter Schmerz und Behinderung leichte Vorteile der Gruppe, welche mit Stabilitätstraining behandelt wurde. Es waren jedoch keine der Resultate statistisch signifikant. Die Resultate der Studie von Rasmussen-Barr et al. (2003) waren bezüglich Schmerz (VAS) und Behinderung (ODI) ebenfalls zu keinem der gemessenen Zeitpunkte signifikant.

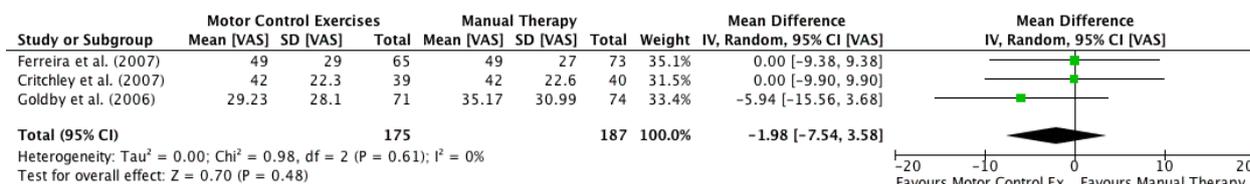
Schmerz (VAS, 0-100) „measured closest to 8 weeks“:



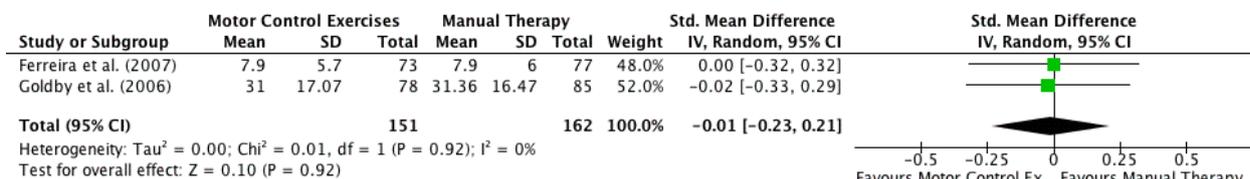
Schmerz (VAS, 0-100) „measured closest to 6 months“:



Schmerz (VAS, 0-100) „measured closest to 12 months“:



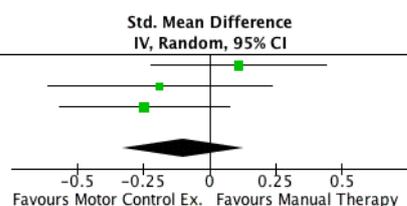
Behinderung (standardisiert) „measured closest to 8 weeks“:



(in Standardabweichungen)

Behinderung (standardisiert) „measured closest to 6 months“:

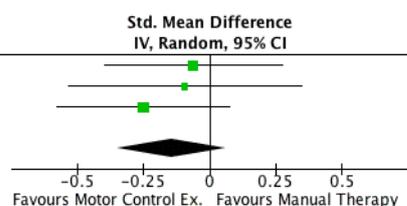
Study or Subgroup	Motor Control Exercises			Manual Therapy			Weight	Std. Mean Difference IV, Random, 95% CI
	Mean	SD	Total	Mean	SD	Total		
Ferreira et al. (2007)	8.4	6.4	68	7.7	6.2	72	36.8%	0.11 [-0.22, 0.44]
Critchley et al. (2007)	7	4.74	41	8	5.64	45	24.7%	-0.19 [-0.61, 0.23]
Goldby et al. (2006)	25.81	17.82	73	30.45	19.67	76	38.5%	-0.25 [-0.57, 0.08]
Total (95% CI)			182			193	100.0%	-0.10 [-0.33, 0.13]
Heterogeneity: Tau ² = 0.01; Chi ² = 2.50, df = 2 (P = 0.29); I ² = 20%								
Test for overall effect: Z = 0.86 (P = 0.39)								



(in Standardabweichungen)

Behinderung (standardisiert) „measured closest to 12 months“:

Study or Subgroup	Motor Control Exercises			Manual Therapy			Weight	Std. Mean Difference IV, Random, 95% CI
	Mean	SD	Total	Mean	SD	Total		
Ferreira et al. (2007)	8.8	6.5	65	9.2	6.6	73	38.2%	-0.06 [-0.40, 0.27]
Critchley et al. (2007)	7.6	5.26	39	8.1	5.32	40	21.9%	-0.09 [-0.53, 0.35]
Goldby et al. (2006)	24.76	17.44	71	29.56	20.52	74	39.9%	-0.25 [-0.58, 0.08]
Total (95% CI)			175			187	100.0%	-0.14 [-0.35, 0.06]
Heterogeneity: Tau ² = 0.00; Chi ² = 0.69, df = 2 (P = 0.71); I ² = 0%								
Test for overall effect: Z = 1.36 (P = 0.17)								



(in Standardabweichungen)

5. Diskussion

5.1. Interpretation der Resultate

Bei den untersuchten Variablen Schmerz und Behinderung konnte zu keinem der untersuchten Zeitpunkte ein statistisch signifikanter Unterschied zwischen Stabilisationstraining und manueller Therapie festgestellt werden. Zusätzlich waren die Unterschiede in den Effekten so klein, dass auch keine klinisch relevanten Unterschiede vorhanden waren. So wurde beim Schmerz die von Ostelo et al. (2008) geforderten 30% Unterschied bei weitem nicht erreicht (kurzfristig: 3.6%, mittelfristig: 5.58%, langfristig: 1.98%).

Auch die Unterschiede in der Variable Behinderung, welche mittels Cohens d standardisiert wurde, waren klein. Cohen bezeichnete einen Effekt mit dem Wert von $d=0.2-0.3$ als klein. Dieser Wert wurde zu keinem der untersuchten Zeitpunkte erreicht (kurzfristig: 0.01, mittelfristig: 0.10, langfristig: 0.14).

Es kann also festgestellt werden, dass auch wenn die Resultate präzise genug wären, so dass die Unterschiede signifikant wären, diese immer noch zu klein sind, um von klinischer Relevanz zu sein.

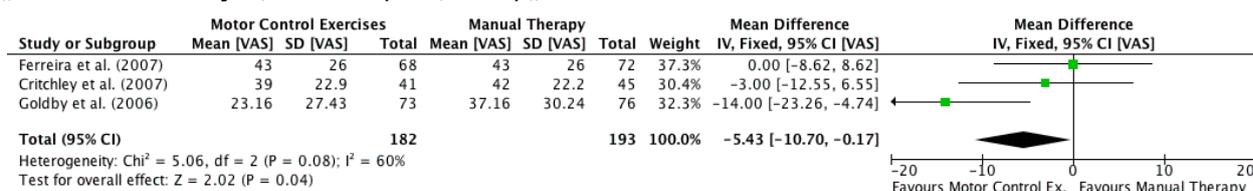
Weiter fällt bei der Betrachtung der Forest-Plots auf, dass Studien mit einer höheren methodologischen Qualität, auch deutlich kleinere Effekte nachgewiesen haben.

5.2. Vergleich mit dem Review von Macedo et al. (2009)

Macedo et al. (2009) hatten zum Ziel, die Wirksamkeit von Stabilitätstraining bei persistierenden lumbalen Rückenschmerzen zu untersuchen. Sie extrahierten dabei die Variablen Schmerz, Behinderung und Lebensqualität; haben die Resultate in eine 0 – 100 Skala umgewandelt und führten anschliessend eine Meta-Analyse durch. Es wurden insgesamt 14 Studien untersucht. Sieben der involvierten Studien verglichen das Stabilisationstraining mit einer minimalen Intervention oder das Stabilisationstraining als Zusatz zu einer herkömmlichen Behandlung. Vier Studien verglichen Stabilisationstraining mit manueller Therapie und fünf Studien Stabilisationstraining mit anderen Trainingsformen. Bezüglich des Vergleiches von

manueller Therapie und Stabilisationstraining konnten Macedo et al. (2009) anhand der gewichteten Zusammenfassung (pooling) der Resultate mittelfristig (>3 Monate und <12 Monate) einen statistisch signifikanten Unterschied zugunsten des Stabilisationstrainings bezüglich Schmerz (gewichtete Mittelwertdifferenz: -5.7, 95% Konfidenzintervall [KI]: -10.7, -0.8) und Behinderung (gewichtete Mittelwertdifferenz: -4.0, 95% KI: -7.6, -0.4) nachweisen. Im Gegensatz dazu konnten in der vorliegenden Studie keine statistisch signifikanten Unterschiede festgestellt werden. Eine Erklärung für diese Differenz ist die unterschiedliche Wahl der Modelle zur Gewichtung der einzelnen Studienresultate in der Zusammenfassung. So verwendeten Macedo et al. das „Fixed-effect“ Modell wenn die Homogenität der untersuchten Studien als genügend beurteilt wurde. Zur Beurteilung der Homogenität wurde dabei I^2 berechnet, wie es von Higgins et al. (2003; zit. nach Borenstein et al., 2009) vorgeschlagen wurde. Allerdings ist dieser Ansatz umstritten, da I^2 nichts über die substantiellen Implikationen der Verteilung aussagt (Borenstein et al., 2009). Vielmehr wird heute empfohlen, die Modell-Wahl nicht auf einen Test der Heterogenität zu basieren. Vielmehr sollte die Modell-Wahl aufgrund des Verständnisses der, den Studien zugrunde liegenden Effekte (oder eben des Effekts) erfolgen (Borenstein et al., 2009). Wie in Kapitel 4.1.5 beschrieben, wurde in dieser Arbeit davon ausgegangen, dass aufgrund der grossen Variabilität in der Patientenzusammensetzung und der Interventionsgestaltung von verschiedenen Effekten ausgegangen werden muss. Eine Analyse der Daten dieser Arbeit mit dem „Fixed-effects“ zeigte, dass dabei bezüglich des Schmerzes nach sechs Monaten signifikante Unterschiede zwischen den Gruppen nachgewiesen werden können (siehe Forest-Plot unten). Diese Differenz zum „Random-effects“ Modell ist durch die im „Fixed-effect“ Modell leicht unterschiedliche Gewichtung der einzelnen Studien erklärbar.

„Fixed-effect“ Analyse, Schmerz (VAS, 0-100) „measured closest to 6 months“:



5.3. Potentieller Bias im Review-Prozess

5.3.1. "Publication Bias"

Ein Umstand, der bei der Erstellung von Reviews zur Besorgnis Anlass geben kann, ist der sogenannte „Publication Bias“. Es gibt Hinweise dafür, dass Studien, welche signifikante Effekte nachweisen konnten, eher publiziert werden, als solche die dies nicht konnten (Dickersin & Min, 1993). Systematische Reviews laufen somit Gefahr, fälschlicherweise signifikante Effekte zu finden. Der „Publication Bias“ kann im vorliegenden Fall allerdings als klein betrachtet werden. Die Erstellung von kontrollierten und randomisierten Studien von physiotherapeutischen Interventionen ist mit einem sehr grossen Aufwand verbunden. Die Wahrscheinlichkeit, dass die Resultate nicht publiziert werden, ist entsprechend klein.

5.3.2. "Measurement Bias" (Datensammlung)

Die methodischen Richtlinien zur Erstellung von systematischen Reviews der Cochrane Back Review Group (Furlan et al., 2009) lieferten Empfehlungen, welche die Datensammlung, das heisst die Extraktion von Studienresultaten zur Analyse, betrifft. So wird empfohlen, diese von mindestens zwei unabhängig von einander arbeitenden Autoren durchführen zu lassen. Dies war im Rahmen dieser Arbeit nicht möglich. Die korrekte Übernahme der Daten wurde jedoch wiederholt und an verschiedenen Tagen überprüft.

5.4. Heterogenität der Behandlungsschemas

5.4.1. Stabilisationstraining

Es existieren bis jetzt noch keine einheitlichen Empfehlungen, welche die Behandlung mittels Stabilitätstraining standardisieren würden. Dieser Umstand wird durch die grosse Variation in den Studien bezüglich der Umsetzung und Dauer der Übungen wiedergegeben. Es ist also davon auszugehen, dass die erzielten Effekte unterschiedlich sind. Zwar wurde dieser Sachlage durch die Wahl des „Random-effects“ Modells entsprochen, trotzdem wird so die Standardabweichung der Verteilung der wirklichen Effekte aller Studien τ zunehmend gross. Das Problem dabei ist, dass bei grossen Standardabweichungen der Nachweis eines signifikanten Unterschiedes zu einem anderen Effekt, z. B. dem von manueller Therapie oder

einem allgemeinen Training, sehr schwierig wird. Dabei bleibt jedoch immer noch das Problem bestehen, dass ein statistisch signifikanter Unterschied nicht von klinischer Relevanz sein muss.

5.4.2. Gruppentherapie versus Einzeltherapie

In den untersuchten Studien fanden die Therapien in den Interventions- und Kontrollgruppen zum Teil in unterschiedlicher Form statt. Während die manuelle Therapie naturgemäss jeweils in Einzelsitzungen durchgeführt wurde, war dies beim Stabilisationstraining nicht immer der Fall. So wurde das Stabilisationstraining bei Goldby et al. (2006) vollkommen und bei Critchley et al. (2007) teilweise in Form einer Gruppentherapie abgehalten. Dieser Kontextfaktor könnte jedoch Einfluss auf den Genesungsverlauf haben. So haben Di Blasi, Harkness, Edzard, Georgiou, & Kleijnen (2001) gezeigt, dass die Art der Beziehung von Patient und Behandelndem (untersucht wurden Ärzte) einen Einfluss auf die Effektivität der Therapie hat. In einer Gruppentherapie ist der Kontakt von Patient und Therapeut viel weniger eng und somit der Aufbau einer fundierten Beziehung schwierig. Es ist nicht auszuschliessen, dass durch die sozialen Kontakte in einer Gruppentherapie ebenfalls positive Effekte, welche den Genesungsverlauf betreffen, auftreten könnten. Dennoch bleibt die Vergleichbarkeit von Einzel- und Gruppentherapie in der Kontroll- und Interventionsgruppe problematisch.

5.5. “Pre-treatment differences”

„Pre-treatment differences“ sind Unterschiede bezüglich der untersuchten Variablen (hier Schmerz und Behinderung), welche schon vor den Interventionen zwischen den verglichenen Gruppen bestanden. Diese sollten sich gemäss (Higgins & Green, 2009) in einem Review von mehreren Studien ausgleichen. Dieser Ausgleich wurde in dieser Arbeit nicht überprüft oder untersucht.

5.6. „Random-effects“ Modell bei einer kleinen Anzahl Studien

Die Auswahl an Studien, zu der in dieser Arbeit gewählten Fragestellung, ist klein. Wie in den Methoden (4.1.5) beschrieben, wurde in der Meta-Analyse das „Random-effects“ Modell gewählt. Ziel dabei ist es den Mittelwert der Effekte μ und deren Standardabweichung τ abzuschätzen (Die Standardabweichung der Verteilung aller Effekte). Bei einer so geringen Anzahl von Studien und somit Effekten wird die

Präzision klein und somit τ gross. Das „Random-effects“ Modell ist dabei immer noch die einzig richtige Wahl, jedoch besteht schlicht ein Informationsdefizit, das eine präzisere Berechnung nicht zulässt. Es werden verschiedene Lösungen dieses Problems angeboten (Borenstein et al., 2009).

1. Eine Variante ist, auf eine Zusammenfassung zu verzichten und alle Effekte nur einzeln aufzulisten. Diese Form birgt allerdings die Gefahr, dass der Leser auf eine Methode zur Zusammenfassung zurückgreift, die in der Literatur häufig als „Vote-Counting“ beschrieben wird. Dabei zählt der Leser quasi alle Studien, die einen signifikanten p-Wert aufweisen und vergleicht diese mit der Anzahl Studien ohne signifikanten p-Wert. Die Differenz sagt jedoch nichts darüber aus, ob ein Effekt tatsächlich vorliegt. Im Gegenteil konnte sogar nachgewiesen werden, dass die Teststärke (Power) des „Vote-Countings“, als statistische Grundlage zur Entscheidungsfindung, kleiner sein kann, als die Teststärke der involvierten Studien. Paradoxerweise kann die Teststärke des „Vote-Countings“ sogar gegen Null tendieren, wenn die Anzahl zusammengefasster Studien grösser wird (Hedges & Olkin, 1980). Erklären lässt sich dieser Umstand zum Teil damit, dass die Abwesenheit eines statistisch signifikanten Effektes keine Evidenz dafür darstellt, dass kein Effekt vorhanden ist.
2. Als weitere Variante könnte eine „Fixed-effect“ Analyse vorgenommen werden. Hierbei ergibt sich unter anderem der Nachteil der Anwendung der Resultate auf grössere Populationen.
3. Ein letzter, jedoch umstrittener Ansatz, wäre die bayessche Interpretation der Daten. Dabei fliessen in die Schätzung der Standardabweichung τ auch Vorinformationen (a priori) ein, die nicht aus den Studien stammen. Die Anwendung dieses Konzeptes braucht jedoch viel Erfahrung und ist bezüglich des philosophischen Ansatzes umstritten.

5.7. Wichtigkeit von Meta-Analysen

Der Wissenszuwachs in der Medizin ist enorm. Wurden im Jahr 1998 noch 467'475 Artikel zur PubMed Datenbank hinzugefügt, so verdoppelte sich diese Zahl innerhalb von nur zehn Jahren nahezu (2008: 816'624 - Corlan, 2009). Um den optimalen Nutzen aus dieser Fülle an Informationen zu ziehen ist eine Form der Zusammenfassung unumgänglich. Wie im Theorieteil gezeigt wurde, haben narrative Reviews den Nachteil der Subjektivität und bei einer Auflistung der Studien zu einer Fragestellung, besteht stets die Gefahr, eines Vote-Countings mit den beschriebenen Unzulänglichkeiten (Kapitel 5.6).

Systematische Reviews sind eine gute Möglichkeit, ohne die oben beschriebenen Nachteile, zu einer besseren Übersicht zu gelangen. Wenn in einer Publikation die Mittelwerte und Standardabweichungen nicht genannt werden, erschwert dies die Erstellung einer Meta-Analyse erheblich. Zwar mag es im Einzelfall sinnvoll sein, den Median und Quantile anzugeben (etwa bei nicht normal-verteilten Daten) aber dennoch sollten die Mittelwerte und Standardabweichungen nicht ausgelassen werden.

6. Schlussfolgerung

Was die Parameter Schmerz und Behinderung betrifft, konnte anhand der untersuchten Studien kein statistisch signifikanter Unterschied der Wirksamkeit bei der Behandlung von lumbalen Rückenschmerzen zwischen Stabilitätstraining und manueller Therapie nachgewiesen werden. Dies heisst jedoch nicht, dass kein Effekt vorhanden ist, sondern, dass ein solcher mit den vorliegenden Daten nicht beweisbar ist.

7. Literaturverzeichnis

- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. West Sussex: Wiley.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, New Jersey, USA: Lawrence Erlbaum Associates.
- Corlan, A. D. (2009). *Medline Trend* [On-Line]. Available: <http://dan.corlan.net/medline-trend.html> (22.8.2010).
- Critchley, D. J., Ratcliffe, J., Noonan, S., Jones, R. H., & Hurley, M. V. (2007). Effectiveness and Cost-Effectiveness of Three Types of Physiotherapy Used to Reduce Chronic Low Back Pain Disability. *Spine*, 14, 1474-1481.
- Di Blasi, Z., Harkness, E., Edzard, E., Georgiou, A., & Kleijnen, J. (2001). Influence of context effects on health outcomes: a systematic review. *The Lancet*, 357, 757-762.
- Dickersin, K., & Min, Y. (1993). Publication Bias: The Problem That Won't Go Away. *Annals of the New York Academy of Sciences*, 703, 135-148.
- Ferreira, M. L., Ferreira, P. H., Latimer, J., Herbert, R. D., Hodges, P. W., Jennings, M. D., et al. (2007). Comparison of general exercise, motor control exercise and spinal manipulative therapy for chronic low back pain: A randomized trial. *Pain*, 131, 31-37.
- Ferreira, P. H., Ferreira, M. L., Maher, C. G., Herbert, R. D., & Refshauge, K. (2006). Specific stabilisation exercise for spinal and pelvic pain: A systematic review. *Australian Journal of Physiotherapy*, 52, 79-88.
- Furlan, A. D., Pennick, V., Bombardier, C., & van Tulder, M. (2009). 2009 Updated Method Guidelines for Systematic Reviews in the Cochrane Back Review Group. *Spine*, 34 (18), 1929-1941.
- Göbel, H. (2001). Epidemiologie und Kosten chronischer Schmerzen. *Schmerz*, 15, 92-98.
- Goldby, L. J., Moore, A. P., Doust, J., & Trew, M. E. (2006). A Randomized Controlled Trial Investigating the Efficiency of Musculoskeletal Physiotherapy on Chronic Low Back Disorder. *Spine*, 31, 1083-1093.

- Hedges, L. V., & Olkin, I. (1980). Vote-Counting Methods in Research Synthesis. *Psychological Bulletin*, 88 (2), 359-369.
- Higgins, J., & Green, S. (2009). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.2 - updated September 2009* [On-Line]. Available: <http://www.cochrane-handbook.org> (20.8.2010).
- Hodges, P., & Richardson, C. (1998). Delayed postural contraction of transversus abdominis in low back pain associated with movement of the lower limb. *Journal of Spinal Disorders*, 11 (1), 46-56.
- Hodges, P., & Richardson, C. (1996). Inefficient muscular stabilisation of the lumbar spine associated with low back pain: a motor control evaluation of transversus abdominis. *Spine*, 21 (22), 2640-2650.
- Kladny, B., Fischer, F. C., & Haase, I. (2003). Wertigkeit der muskulären segmentalen Stabilisierung zur Behandlung von Rückenschmerzen und Bandscheibenerkrankungen im Rahmen der ambulanten Rehabilitation. *Zeitschrift für Orthopädie und ihre Grenzgebiete*, 141, 401-405.
- Läubli, T., & Müller, C. (2009). *Arbeitsbedingungen und Erkrankungen des Bewegungsapparates – Geschätzte Fallzahlen und volkswirtschaftliche Kosten für die Schweiz* [On-Line]. Available: <http://www.seco.admin.ch/dokumentation/publikation/00008/00022/02439/index.html?lang=de> (15. 7.2010).
- Lederman, E. (2009). The myth of core stability. *Journal of Bodywork & Movement Therapies*, 14 (1), 84-98.
- Macedo, L. G., Maher, C. G., Latimer, J., & McAuley, J. H. (2009). Motor Control Exercise for Persistent, Nonspecific Low Back Pain: A Systematic Review. *Physical Therapy*, 89 (1), 9-25.
- Maher, C. G., Sherrington, C., Herbert, R. D., Moseley, A. M., & Elkins, M. (2003). Reliability of the PEDro Scale for Rating Quality of Randomized Controlled Trials. *Physical Therapy*, 83 (8), 713-721.
- Ostelo, R. W., Deyo, R. A., Stratford, P., Waddell, G., Croft, P., Von Korf, M., et al. (2008). Interpreting Change Scores for Pain and Functional Status in Low Back Pain: Towards International Consensus Regarding Minimal Important Change. *Spine*, 33 (1), 90-94.

- Rasmussen-Barr, E., Nilsson-Wikmar, L., & Arvidsson, I. (2003). Stabilizing training compared with manual treatment in sub-acute and chronic low-back pain. *Manual Therapy, 8* (4), 233-241.
- Richardson, C., Jull, G., Hodges, P. W., & Hides, J. (1999). *Therapeutic Exercise for Spinal Segmental Stabilisation in Low Back Pain: Scientific Basis and Clinical Approach*. Edinburgh: Churchill Livingstone.
- The Cochrane Collaboration. (2008). Review Manager (RevMan) [Computer program] Version 5.0.24. Copenhagen: The Nordic Cochrane Center, Denmark.
- Verhagen, A. P., de Vet, H. C., de Bie, R. A., Kessels, A. G., Boers, M., Bouter, L. M., et al. (1998). The Delphi List: A Criteria List for Quality Assessment of Randomized Clinical Trials for Conducting Systematic Reviews Developed by Delphi Consensus. *Journal of Clinical Epidemiology, 51*, 1235-1241.

8. Danksagung

Ich möchte mich herzlich bei Herrn Dr. Jan Kool für die Betreuung dieser Bachelor Arbeit, das fortwährende Interesse und die ständige Ansprechbarkeit bedanken.

Weiter gilt mein Dank Frau Katja Schulze für die grossartige Unterstützung, das beharrliche Anspornen und das Korrekturlesen.

Ich bedanke mich herzlich bei meinen Eltern für ihre uneingeschränkte Unterstützung.

9. Eigenständigkeitserklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbständig, ohne Mithilfe Dritter und unter Benützung der angegebenen Quellen verfasst habe.

Zürich, 9. September 2010

Pascal Oliver Kaufmann