

Estimating the accuracy of muscle response testing: two randomised-order blinded studies

von Anne M. Jensen 2016

Dt. Fassung DGAK www.dgak.de Juni 2017 (übersetzt von Antonia Bose)

Zusammenfassung

In den vorliegenden Studien wurde untersucht, ob man durch Anwendung des Muskeltests eine Lüge von der Wahrheit unterscheiden kann. Der Patient traf hierzu eine wahre oder unwahre Aussage, woraufhin der Fachausübende den Muskeltest durchführte. Eine schwache Muskelreaktion wies dabei auf eine Lüge, ein starkbleibender Muskel auf eine wahre Aussage hin. Am Ende wurde ausgewertet wie oft der Fachausübende eine Lüge oder eine wahre Aussage durch den Muskeltest korrekt identifizierte. Verglichen wurde er außerdem mit der Zufallswahrscheinlichkeit von 50% (Lüge/Wahrheit) und einigen Versuchsblöcken, in denen der Fachausübende den Muskeltest nicht ausführte, sondern sich auf seine Intuition verließ und so riet, ob der Patient die Wahrheit sagte oder nicht. Schlussfolgernd lässt sich sagen, dass der Muskeltest akkurater zwischen einer Lüge und der Wahrheit unterscheiden konnte als Intuition. Auch lag seine Aussagekraft deutlich höher als 50%.

Grundlagen

Der kinesiologische, manuelle Muskeltest (MMT) ist eine nicht-invasive Methode, welche oft im Gesundheitswesen verwendet wird. So wird er beispielsweise von Physiotherapeuten, Chiropraktikern, Osteopathen und Ärzten für die Untersuchung von neuromuskulärer Funktionalität angewandt. Eine Art des MMTs stellt der Muskelreaktionstest (MRT) dar, durch welchen Muskeln getestet werden, nicht etwa um muskuläre Kraft, sondern neuronale Kontrolle, zu bewerten. Der MRT wurde in den 1970ern und 1980ern durch die Arbeit von Goodheart u.a. ins Leben gerufen. Da der MRT weltweit schätzungsweise von über einer Millionen Menschen genutzt wird, ist es wichtig seine Validität zu untersuchen. Was den MRT von anderen manuellen Muskeltests unterscheidet ist, dass nur ein Muskel getestet wird. Dieser wird mehrfach hintereinander getestet um festzustellen, ob bestimmte Beschwerden vorliegen, welche

behandelt werden müssen. Hierzu zählen beispielsweise Schmerzen im unteren Rücken, Phobien und Lebensmittelallergien.

Eine etablierte Anwendungsmöglichkeit des MRTs innerhalb einer Therapiesitzung ist es, Patientenreaktionen auf semantische Stimuli (z.B. gesprochene Aussagen) zu bewerten. Der semantische Stimulus kann sowohl vom Klienten selber, als auch vom Fachausübenden ausgesprochen werden. Der Praktizierende zeichnet dann die muskuläre Resistenz des Patienten auf. Hierzu wird ein Druck auf den jeweiligen Muskel ausgeübt, während gleichzeitig die Aussage verbalisiert wird.

Eine frühere Studie mit 89 Probanden zeigte, dass wahre Aussagen zu weitaus mehr muskulärer Resistenz führten als unwahre Aussagen. Einige wichtige Details wurden in diesem Bericht jedoch ausgelassen. Hierzu zählten die Anzahl der Fachausübenden und, vor allem, in wie weit die Fachausübenden und die Patienten von der Hypothese wussten, d.h. experimentell blind waren. Im Jahre 2009 wurde das Protokoll einer randomisierten, kontrollierten Studie veröffentlicht, in welchem der MRT im Rahmen einer Therapie verwendet wurde. Die dazugehörigen Ergebnisse wurden noch nicht veröffentlicht.

Unser Ziel war es zu beurteilen mit welcher Genauigkeit der MRT wahren von unwahren Aussagen unterscheiden kann. Dies wurde unter randomisierten, kontrollierten Bedingungen untersucht. Ein weiteres Ziel war der Vergleich zwischen der Treffgenauigkeit des MRTs auf der einen Seite und der Treffgenauigkeit eines Fachausübenden, welcher sich lediglich auf seine Intuition verlässt, auf der anderen Seite.

Studiendesign

Bei diesen Studien handelte es sich um zwei prospektive Studien über diagnostische Testgenauigkeit. Sie wurden anhand von zwei klinischen Versuchsregistrierungen (Australian New Zealand Clinical Trial Registry (ANZCTR; www.anzctr.org.au; ID # ACTRN126090000455268) und US-based ClinicalTrials.gov (ID # NCT010663612) verzeichnet. Die Genehmigung für die Durchführung kam aus dem Vereinigten Königreich (Oxford Tropical Research Ethics Committee) und den USA (Parker University Institutional Review Board). Alle Beteiligten willigten der Veröffentlichung ein. Alle Grundsätze der Erklärung von Helsinki wurden eingehalten. Ergebnisse wurden gemäß STARD (Standards for the Reporting of Diagnostic Test Accuracy Studies) Richtlinien analysiert und dargestellt.

Das Modell, welches in der vorliegenden Studie untersucht wurde, lautete wie folgt: Lügen (d.h. eine unwahre Aussage treffen) führt zu einer Muskelabschwächung, während das Verbalisieren einer Aussage, die der Wahrheit entspricht, zu einer starken MRT-Reaktion führt. Eine schwache muskuläre Reaktion wird von uns als ein diagnostischer Indikator für eine Lüge angesehen. Ein starkbleibender Muskel hingegen wird als Hinweis auf den hohen Wahrheitsgehalt der Aussage betrachtet.

Zum Vergleich berücksichtigten wir noch einen zweiten Indikator: Intuition. Hierfür wurden die Fachausübenden gebeten, sich auf ihre Intuition zu verlassen um den Wahrheitsgehalt einer Aussage zu bestimmen. Sie benutzten dafür keinen MRT. Da Lügen bekannterweise eine Reihe physiologischer Merkmale mit sich ziehen, wurden die Praktizierenden gebeten, nur ihre Sinne zu gebrauchen um einer Lüge auf die Spur zu kommen: Sehen (z.B. um Körpersprache und

Gesichtsausdrücke zu betrachten), Hören (z.B. Veränderungen in der Stimme) und Fühlen (z.B. Veränderungen bezüglich der Hauttemperatur).

In beiden Experimenten wechselten sich vier Blöcke á zehn MRTs mit vier Blöcken á zehn Experimenten mit Intuition ab. Es wurde immer mit einem MRT-Block begonnen. Die Praktizierenden alleine bestimmten die Resultate der MRTs und Intuitionen und es waren auch nur sie alleine, die die Resultate in einen Computer mithilfe einer Tastatur eingaben.

Experiment 1

Probanden

Es gab zwei Teilnehmergruppen: (1) Fachausübende aus dem Gesundheitssektor („Fachausübende“; n=48), welche regelmäßig den MRT praktizierten, und (2) Testpatienten („TPs“; n=48). Letztere waren mit dem MRT nicht vertraut. 48 Pärchen aus je einem Fachausübenden und einem TP wurden gebildet. Alle 96 Teilnehmer wurden anhand direktem Kontakt (E-Mail oder Telefon), Social Media oder Mundpropaganda geworben. Die Freiwilligen hatten folgende Kriterien zu erfüllen, um zu diesem Experiment zugelassen zu werden: Sie mussten 18-65 Jahre alt sein, komplett funktionsfähige und schmerzfreie obere Extremitäten besitzen, und die englische Sprache fließend beherrschen. Sie wurden von der Studie hingegen ausgeschlossen, wenn sie blind, taub oder stumm waren. Keiner der TPs kannte seinen zugeteilten Fachausübenden. Alle Fachausübenden, die den Wunsch hatten teilzunehmen und die Kriterien erfüllten, wurden zugelassen. Hierbei spielte es keine Rolle, welchen Beruf sie ausübten, welche MRT-Techniken sie beherrschten, oder inwieweit sie bereits

Erfahrung mit dem MRT gesammelt hatten. Es wurden keine Tests durchgeführt, um festzustellen inwiefern die Praktizierenden den Muskeltest beherrschten.

Test des Hauptindex: MRT

Während eines MRT wird externe Kraft auf eine Gliedmaße ausgeübt, wodurch Widerstand in einem bestimmten Muskel entsteht. Zunächst hält der Patient ein bestimmtes Gelenk in einer festen Position – meist leicht gebeugt. Dann übt der Fachausübende Druck aus. Dies passiert üblicherweise über einen längeren Zeitraum hinweg, da der Patient versucht diesem Druck durch isometrische Kontraktion entgegenzuwirken. Beispielweise hält der Fachausübende die Schulter des Patienten (Glenohumeralgelenk) in einem 90-Grad-Winkel. Hierbei zeigt die Handinnenfläche nach unten, während der Fachausübende den anterioren Deltamuskel testet. Er entscheidet dann subjektiv, ob der Muskel schwach wurde oder stark geblieben ist.

Fachausübende können sich in der Stärke des Druckes und der genauen Platzierung ihrer Hand unterscheiden. Die Hand liegt normalerweise auf dem unteren Unterarm des Patienten – nahe des Handgelenks. Für das vorliegende Experiment wurden die Fachausübenden gebeten, ihren üblichen Praktiken nachzukommen.

Testmethoden

TPs trafen insgesamt 40 wahre und unwahre Aussagen wie folgt: Ihnen wurden Bilder auf einem Computerbildschirm gezeigt, welcher sich außerhalb des Sichtfeldes des Fachausübenden befand. Während dem TP ein vom Computer zufällig ausgewähltes Bild gezeigt wurde, bekam er durch Instruktionen durch

einen Knopf im Ohr. Die Instruktionen wurden von einer Computerstimme gesprochen und waren für den Praktizierenden nicht hörbar. Sie lauteten wie folgt: „Ich sehe ein(e/n) ...“. Ob die Aussage der Wahrheit entsprach, wurde ebenfalls zufällig von der Software (DirectRT Research Software) ausgewählt. Lügen kamen in etwa 50% ($\pm 3\%$) der Fälle vor. Dem Fachausübenden wurde ein anderer Bildschirm gezeigt. Er sah entweder das gleiche Bild wie der TP (d.h. nicht-blinde Studie) oder einen komplett schwarzen Bildschirm (d.h. blinde Studie). Die TPs wussten weder etwas über das Ziel der Studie, noch wussten sie in welcher Proportion die wahren und falschen Aussagen auftauchten und welchen Bildschirm der Fachausübende sah. Neutrale Bilder (d.h. emotional neutral) wurden aus dem International Affective Picture System (IAPS) ausgewählt und mit neutralen Wörtern gepaart, welche anhand der Affective Norms for English Words (ANEW) selektiert wurden.

Nach jeder Aussage des TPs wurde der Fachausübende gebeten, die Wahrheit der Aussage festzustellen: zehn Mal anhand des MRTs, dann zehn Mal anhand reiner Intuition. Dies wurde in Zehner-Blöcken fortgeführt. Die Fachausübenden gaben ihre Wertung über eine Tastatur ein, indem sie eine bestimmte Taste drückten. Diese Tastatur war mit dem Laborcomputer verbunden, welcher automatisch die Ergebnisse generierte. Bevor das eigentliche Experiment anging, wurde den Teilnehmer und den Fachausübenden Zeit gegeben, sich mit der Prozedur des Experiments vertraut zu machen. Während der Datensammlung war der zuständige Forscher anwesend, jedoch griff er nicht ins Geschehen ein.

Zudem wurden die TPs gebeten, zwei kurze Fragebögen auszufüllen; einen vor Beginn des Experiments und einen nachdem das Experiment abgeschlossen war. Der erste TP-Fragebogen enthielt Fragen bezüglich des Alters, Geschlechts,

Händigkeit, MRT-Erfahrung, Sicherheit im Umgang mit dem MRT, und Vertrauen in den Fachausübenden und seine MRT-Fähigkeit. Der erste Fragebogen für die Praktizierenden umfasste ebenfalls Fragen bezüglich des Alters, Geschlechts und Händigkeit. Hierzu kamen jedoch noch Fragen zur Praxisart, Praxiserfahrung in Jahren, MRT-Erfahrung in Jahren, Selbsteinschätzung über MRT-Kompetenz, welche spezifischen MRT-Techniken er oder sie besaß, und wie groß das Vertrauen in den MRT generell und die eigenen Fähigkeiten war. Vertrauenslevel wurden anhand einer 10cm VAS (Visual Analogue Scale) gemessen. Hierbei stand das linke Ende für „kein Vertrauen“ und das rechte Ende für „komplettes Vertrauen“. Alle Teilnehmer wurden gebeten ihr Level mit einem Strich zu markieren, woraus zu einem späteren Zeitpunkt ein Zahlenwert ermittelt wurde. Selbsteinschätzungen über MRT-Fähigkeiten wurden anhand einer 5-Punkt Likert-Skala ermittelt, welche von 0 (kein) bis 4 (Experte) reichte. Jahresangaben (z.B. Altersangaben, Erfahrungsjahre, etc.) wurden als Laufvariablen, und andere Variablen (z.B. Geschlecht, Beruf, MRT-Techniken, etc.) als kategoriale Variablen kodiert.

In den Fragebögen, die nach dem Experiment ausgefüllt wurden, wurden die Teilnehmer wieder gebeten ihre Zuversicht zu beurteilen. Zudem wurden TPs gefragt, ob ihnen etwas während des Experiments auffiel. Hierzu konnten sie mit einer offenen Aussage antworten. So wurde festgestellt, ob sie etwas über die Hypothese der Studie (d.h. eine Lüge führt zu einer „schwachen“ MRT-Reaktion) wussten. Um die Genauigkeit des Experiments sicherzustellen, war der zuständige Forscher zu jeder Zeit des Experiments anwesend.

Experiment 2

Probanden

Die Teilnehmer wurden ähnlich wie im ersten Experiment ausgewählt, jedoch wurde die Teilnehmeranzahl auf 20 Paare reduziert. TPs, welche bereits mit dem MRT vertraut waren, wurden diesmal auch miteinbezogen. Außerdem wurden TPs und Fachausübende zusammengetan, auch wenn sie sich bereits vorher kannten.

Testmethoden

Die Testmethoden dieses Experiments glichen denen des ersten, mit folgenden Ausnahmen: 1) Keiner der Fachausübenden wusste, ob der TP die Wahrheit sagt oder nicht; 2) Die TP-Fachausübenden-TP-Paare wurden für das komplette Experiment alleine gelassen; 3) Vor dem MRT bewerteten die Fachausübenden ihren aktuellen Angstzustand; und 4) Genau 50% der Aussagen waren Lügen.

Statistische Methoden

Um die Treffgenauigkeit des MRTs eines jeden Fachausübenden-TP-Paares zu untersuchen, betrachtete man die Fälle, in denen folgende zwei Kriterien erfüllt wurden: der MRT musste korrekt identifizieren, ob es sich um eine Lüge oder eine wahre Aussage handelte, und es musste eine blinde Studie sein. Für das erste Experiment wurden Probedaten verwendet, um die Teilnehmerzahl festzulegen. In den Probedurchläufen enthielt man eine MRT-Genauigkeit von 67.7% (95% Konfidenzintervall 52.6% bis 82.8%). Von diesen Daten ausgehend wurde die angemessene Paaranzahl auf 48 geschätzt. Mit diesen 48 Paaren sollte es folglich möglich sein zu untersuchen, ob der MRT eine Lüge von einer Wahrheit unterscheiden kann. Es wurde von einem 95% Konfidenzintervall und

80% Power ausgegangen. Aus diesem Grund geben wir im Folgenden die Durchschnittsgenauigkeit des MRTs mit einem 95% Konfidenzintervall an.

Intuitions-genauigkeit wurde ähnlich definiert und dargestellt. Vor der Analyse wurden Annahmen bezüglich einer Normalverteilung graphisch überprüft. Die Nullhypothese, dass die Durchschnittsdifferenz zwischen Genauigkeit des MRTs und Genauigkeit von Intuition gleich null ist wurde anhand abhängiger t-Tests untersucht. Sekundäre Ergebnisse wie Sensitivität, Spezifität, positive prädiktive Werte und negative prädiktive Werte wurden separat analysiert und ausgewiesen. Um Zusammenhänge zwischen Genauigkeit und Kovariaten (Alter, Geschlecht, Beruf, Praxiserfahrung in Jahren, aktueller Praxisstatus, Dauer und Ausmaß von MRT-Erfahrung, gelernte MRT-Techniken, Händigkeit, selbsteingeschätzte Zuversicht bezüglich MRT-Durchführung, und selbsteingeschätzter Grad von Testangst) zu überprüfen wurde eine lineare Regression angewendet. Alle Analysen basierten lediglich auf jenen Experimenten, welche blind durchgeführt wurden.

Ergebnisse

Experiment 1

Probanden

Zwischen Juni 2010 und Oktober 2011 meldeten sich achtundvierzig Paare zum Experiment in den USA und in der UK an. Sechs freiwillige Fachausübende mussten aus der Studie ausgeschlossen werden; vier aufgrund ihres zu hohen Alters (>65 Jahre), einer aufgrund Sprachschwierigkeiten im Englischen, und ein weiterer aufgrund beeinträchtigtem Hörvermögen. Unter den 48 TPs waren 31 Frauen und 17 Männer. Das Durchschnittsalter betrug 39.0 Jahre

(Standardabweichung: 11.4). Die 48 Fachausübende spalteten sich in 32 Frauen und 16 Männer auf. Unter ihnen betrug das Durchschnittsalter 49.3 Jahre (Standardabweichung: 12.0). Der Mittelwert der Praxiserfahrung lag bei 11.5 Jahren (Interquartilsabstand: 7.3 bis 20.8). Der Mittelwert der MRT-Erfahrung betrug ebenfalls 11.5 Jahre (Interquartilsabstand: 5.3 bis 17.3), während der Mittelwert der MRT-Ausübung pro Tag bei 2.9 Stunden (Interquartilsabstand 1.0 bis 6.0) lag. Auf einer Skala von 0 bis 4 betrug der Durchschnitt der Selbsteinschätzung bezüglich MRT-Kompetenz 3.1 (Standardabweichung: 0.2).

Testergebnisse

Diagnostische MRT-Treffgenauigkeit (d.h. richtige Identifizierung einer Lüge bzw. Wahrheit) reichte in den blinden Experimenten von 0.4 bis 0.917. Der Durchschnitt lag bei 0.659 (95% Konfidenzintervall: 0.623 bis 0.695). Wenn die Fachausübenden rein intuitiv versuchten Lügen zu identifizieren, reichte die Treffgenauigkeit von 0.238 bis 0.636 mit einem Durchschnittswert von 0.474 (95% Konfidenzintervall: 0.449 bis 0.500). Der Durchschnittswert der MRT-Treffgenauigkeit war signifikant höher als der der intuitiven Treffgenauigkeit ($p=0.01$). Es konnte kein signifikanter Zusammenhang zwischen der Treffgenauigkeit beim MRT und der intuitiven Treffgenauigkeit eines Fachausübenden gefunden werden ($r=-0.03$, $p=0.86$, 95% Konfidenzintervall: -0.31 bis 0.26).

Die Durchschnittsensitivität des MRTs (d.h. seine Fähigkeit eine Lüge korrekt zu identifizieren) betrug 0.568 (95% Konfidenzintervall: 0.504 bis 0.633), während die Durchschnittsspezifität (d.h. Genauigkeit bei der Identifizierung einer Lüge) bei 0.734 (95% Konfidenzintervall: 0.687 bis 0.782) lag. Die

Durchschnittswerte der positiven prädiktiven Werte und negativen prädiktiven Werte des MRTs betragen jeweils 0.663 (95% Konfidenzintervall: 0.607 bis 0.718) und 0.667 (95% Konfidenzintervall: 0.625 bis 0.708).

Des Weiteren wurden verschiedene Analysen aufgrund unterschiedlicher Arbeitserfahrung durchgeführt. Durchschnittstreffgenauigkeit des MRTs der 20 teilnehmenden Chiropraktikern lag bei 0.670 (95% Konfidenzintervall: 0.611 bis 0.729), während die Nicht-Chiropraktiker einen Durchschnittswert von 0.642 (95% Konfidenzintervall: 0.593 bis 0.691) hatten. Zwischen diesen beiden Gruppen wurde kein signifikanter Unterschied bezüglich MRT-Treffgenauigkeit gefunden ($p=0.45$). Vollzeitpraktizierende ($n=26$) wiesen einen Durchschnittswert von 0.663 (95% Konfidenzintervall: 0.612 bis 0.715) auf, während die Durchschnittstreffgenauigkeit des MRTs bei Halbzeitpraktizierenden ($n=13$) bei 0.682 (95% Konfidenzintervall: 0.618 bis 0.746) lag. Jene Fachausübende, die aktuell nicht praktikzierten ($n=7$) kamen auf einen Durchschnittswert von 0.569 (95% Konfidenzintervall: 0.465 bis 0.673). Zwischen diesen drei Gruppen konnte ebenfalls kein statistisch signifikanter Unterschied festgestellt werden. Fachausübende, welche angaben, Experten des Muskeltests zu sein (Level 4 von 4; $n=15$) wiesen eine Durchschnittstreffgenauigkeit von 0.682 (95% Konfidenzintervall: 0.617 bis 0.747) auf, Fachausübende in der zweithöchsten Kategorie (Level 3 von 4; $n=19$) einen Wert von 0.666 (95% Konfidenzintervall: 0.605 bis 0.728), und jene, die angaben, zu den unteren beiden Kategorien zu gehören (Level 1 oder 2 von 4; $n=12$) kamen auf eine Durchschnittstreffgenauigkeit von 0.600 (95% Konfidenzintervall: 0.5258 bis 0.672). Da $p=0.35$, liegt auch hier kein signifikanter Unterschied vor. Fälle, in denen der TP die Studienhypothese riet ($n=21$) ergaben

einen Durchschnittswert von 0.661 (95% Konfidenzintervall: 0.591 bis 0.730), während für die anderen Paare (d.h. in denen der TP die Hypothese kannte; n=27) ein Wert von 0.649 (95% Konfidenzintervall: 0.610 bis 0.688) entstand. Auch hier liegt kein signifikanter Unterschied vor ($p=0.38$).

Außerdem wurde auch keine besondere Entwicklung der Treffgenauigkeit festgestellt, während das Experiment voranschritt. Eine Post-hoc-Analyse fand zudem keinen Unterschied zwischen Experimenten, welche in einer lauten Umgebung durchgeführt wurden und anderen Umgebungen ($p=0.46$). Mit einer Ausnahme (Ermüdung des Schultermuskels bei n=7 von 96) gab es keine weiteren Komplikationen während des Experimentes.

Experiment 2

Probanden

Zwanzig Fachausübende-TP-Paare meldeten sich zwischen Juli und November 2011 für das Experiment in den USA und der UK an. Unter ihnen waren 13 weibliche und 7 männliche Fachausübende, und 8 weibliche und 12 männliche Testpatienten. Das Durchschnittsalter der Fachausübenden betrug 49.3 (Standardabweichung: 12.0) Jahre, und das der TPs 40.8 (Standardabweichung: 12.8) Jahre. 14 der 20 teilnehmenden Fachausübenden waren Chiropraktiker, 2 Fachausübende, welche im Gebiet psychologischer Krankheiten tätig waren, 1 Akupunkteur, und 3 weitere Fachausübende. Vierzehn Fachausübende waren Vollzeitarbeitende, 4 Halbzeitarbeitende, und 2 weitere Fachausübende waren zu dem Zeitpunkt nicht berufstätig. Der Mittelwert der Praxiserfahrung betrug 18.0 (Interquartilsabweichung: 17.0) Jahre, während der Mittelwert der MRT-Erfahrung bei 14.0 (Interquartilsabweichung: 16.0) Jahren lag. MRT-Ausübung

pro Tag hatte einen Mittelwert von 4.0 (Interquartilsabweichung: 4.0). Die durchschnittliche MRT-Kompetenz, welche von den Teilnehmern selbst eingeschätzt wurde, lag auf einer Skala von 0 bis 4 bei 3.2 (Standardabweichung: 0.7).

Testergebnisse

Im zweiten Experiment lag die durchschnittliche Treffgenauigkeit mit der der MRT eine Lüge von einer Wahrheit unterschied bei 0.594 (95% Konfidenzintervall: 0.541 bis 0.647) und reichte von 0.425 bis 0.825. Die durchschnittliche Genauigkeit von alleiniger Intuition lag bei 0.514 (95% Konfidenzintervall: 0.483 bis 0.544) und reichte von 0.375 bis 0.625. Dies stellt einen signifikanten Unterschied zwischen den beiden Methoden dar ($p=0.01$). Die Wahrscheinlichkeit mithilfe des MRTs eine Lüge auszumachen war nicht nur signifikant höher als die Identifizierung von Lügen durch Intuition, sondern auch größer als Zufallswahrscheinlichkeit (i.e. 50%; $p<0.01$). Es gab keine signifikante Korrelation zwischen der Treffgenauigkeit eines einzelnen Fachausübenden bei der Verwendung des MRTs und der Verwendung von Intuition ($r=0.07$, $p=0.77$, 95% Konfidenzintervall: -0.38 bis 0.50).

Die Durchschnittsensitivität (d.h. mit welcher Genauigkeit der MRT Lügen identifizierte) lag bei 0.583 (95% Konfidenzintervall: 0.534 bis 0.631) und die Durchschnittsspezifität (d.h. mit welcher Genauigkeit der MRT wahre Aussagen identifizierte) bei 0.631 (95% Konfidenzintervall: 0.553 bis 0.673). Der Durchschnitt der positiven prädiktiven Werte des MRTs betrug 0.685 (95% Konfidenzintervall: 0.616 bis 0.754) und der der negativen prädiktiven Werte des MRTs 0.503 (95% Konfidenzintervall: 0.421 bis 0.584).

Die Treffgenauigkeit bei der Verwendung des MRTs durch die 14 Chiropraktiker (0.607; 95% Konfidenzintervall: 0.535 bis 0.679) war nicht signifikant anders ($p=0.36$) im Vergleich zur Treffgenauigkeit der 6 Nicht-Chiropraktikern (0.563; 95% Konfidenzintervall: 0.478 bis 0.647). Vollzeitpraktizierende ($n=14$) wiesen einen Durchschnittswert von 0.561 (95% Konfidenzintervall: 0.504 bis 0.618), Halbzeitpraktizierende ($n=4$) von 0.706 (95% Konfidenzintervall: 0.508 bis 0.905), und aktuell Nicht-Praktizierende ($n=2$) von 0.600 (95% Konfidenzintervall: 0.000 bis 1.000) auf. Es gab keinen signifikanten Unterschied zwischen diesen Gruppen ($p=0.07$) bezüglich MRT-Genauigkeit. Jene Fachausübende, die sich als Experten bezüglich MRT-Kompetenz einschätzten (Level 4 von 4; $n=7$), hatten eine durchschnittliche MRT-Treffgenauigkeit von 0.611 (95% Konfidenzintervall: 0.470 bis 0.751) und diejenigen, die sich in die zweithöchste Kategorie einordneten (Level 3 von 4; $n=10$) wiesen einen Durchschnittswert von 0.590 (95% Konfidenzintervall: 0.518 bis 0.662) auf. Die weiteren 3 Fachausübenden gaben an, zu den unteren beiden Kategorien (Level 1 oder 2 von 4; $n=3$) zu gehören und hatten eine durchschnittliche Treffgenauigkeit von 0.567 (95% Konfidenzintervall: 0.387 bis 0.746). Zwischen diesen Gruppen gab es keinen signifikanten Unterschied ($p=0.86$) bezüglich MRT-Treffgenauigkeit. TPs, die die Hypothese rieten ($n=6$) hatten eine durchschnittliche MRT-Genauigkeit von 0.621 (95% Konfidenzintervall: 0.507 bis 0.735), während jene, die die Hypothese kannten ($n=14$) durchschnittlich 0.582 (95% Konfidenzintervall: 0.515 bis 0.650) aufwies. Zwischen diesen beiden Gruppen wurde ebenfalls kein statistisch signifikanter Unterschied gefunden ($p=0.49$). Ähnlich wie im ersten Experiment kamen auch in diesem keine Komplikationen, mit Ausnahme von Ermüdung der Muskeln ($n=4$ von 40), auf.

Diskussion

Generelle Befunde

Bei der Unterscheidung zwischen wahren und unwahren Aussagen führte die Verwendung des Muskeltests kontinuierlich zu besseren Resultaten als man durch Zufall erwartet hätte. Zudem erzielte die Verwendung des MRTs höhere Ergebnisse als wenn der selbe Fachausübende sich lediglich auf seine Intuition verlies. Dies deutet darauf hin, dass der Erfolg des MRTs auf den Test selber zurückzuführen ist und nicht beispielsweise auf die Fähigkeit Körpersprache oder Tonlagen korrekt zu deuten. Diese Studien stellen einen wichtigen Schritt in Richtung Akzeptanz des MRTs als diagnostisches Instrument dar. Außerdem wird gezeigt, dass wissenschaftliche Methoden, wie experimentelle Blindheit und Randomisierung, auch genutzt werden können, um den wissenschaftlichen Wert von komplementären und alternativ-medizinischen Tests zu erfassen.

Alle Analysen, welche zuvor präsentiert wurden, basierten auf die Auswertung blinder Experimente. Ergebnisse von nicht-blinden Experimenten und Experimenten, in welchen der Fachausübende aktiv getäuscht wurde, wurden hier nicht genannt.

Stärken und Einschränkungen

In den vorliegenden Studien wurden die MRT-Methoden nicht standardisiert. Es hätte beispielsweise ein Presskolben verwendet werden können, um Druck zu überprüfen. Allerdings wurde so gewährleistet, dass die Fachausübenden den MRT so ausführten, wie sie es auch im richtigen Leben getan hätten. Die Entscheidung, keine Presskolben zu verwenden wurde von vorherigen Studien

unterstützt. Diese fanden heraus, dass durch die Verwendung von Presskolben deutliche Unterschiede zwischen „starken“ und „schwachen“ Muskeln aufgezeigt wurden, wodurch die Nutzung dieser überflüssig wurde. Andere Stärken der vorliegenden Studien sind beispielsweise die experimentelle Blindheit und sowohl klar definierte Vergleichsnormen, als auch klar definierte experimentelle Konditionen. Jedoch standen die Aussagen, welche als Standards verwendet wurden, in keinem Bezug zum eigentlichen Ziel der Studie. Zudem wurde der MRT nicht mit anderen lügenerkennenden Methoden, wie zum Beispiel dem Polygraphen, verglichen. Andere mögliche Anwendungsgebiete des MRTs, wie beispielsweise Feststellung von Lebensmittelallergien oder notwendigen Nahrungsergänzungsmitteln, oder die Bewertung von sportlichen Leistungen, überschreiten das Ausmaß dieser Studien.

Obwohl die Fachausübenden die Wahrheit der Aussagen nicht kannten, so waren die Testpatienten nicht notwendigerweise experimentell blind. Da jedoch keine signifikanten Unterschiede zwischen TPs, welche die Hypothese der Studie kannten, und TPs, welche die Hypothese nicht kannten, aufkamen, scheint die experimentelle Blindheit der TPs keine Auswirkungen gehabt zu haben. Methodisch hätten die Studien gestärkt werden können, indem man die Reihenfolge der Blöcke (d.h. ob ein Paar mit MRT oder Intuition anfängt) randomisiert hätte.

Stärken und Schwächen im Vergleich mit anderen Studien

Es gab bereits eine weitere Studie, die versuchte, die Treffgenauigkeit des MRTs in Bezug auf Lügen zu untersuchen. Jedoch wurden dabei einige Eigenschaften der Fachausübenden vorenthalten: man wusste nicht, wie viele Fachausübende

teilnahmen, wie sie angeworben wurden, welche Kriterien sie zu erfüllen hatten, um zur Studie zugelassen zu werden, und inwiefern sie blind waren. Da es sich hierbei um wichtige Informationen handelt, wird so die Brauchbarkeit der Resultate eingeschränkt. Eine andere Studie untersuchte die Fähigkeit von Fachausübenden, eine starke von einer schwachen Reaktion zu unterscheiden. Da hier die Verbindung zu wahren bzw. unwahren Aussagen und deren Identifizierung fehlt, ist ein Vergleich eher schwierig.

Konsequenzen für die klinische Anwendung und zukünftige Forschung

Mit der vorliegenden Studie konnten wir zeigen, dass der MRT besser wahren von unwahren Aussagen unterscheiden kann, als durch puren Zufall möglich wäre. Die Aussagen im Experiment entsprachen jedoch nicht klinisch-relevanten Aussagen und durchschnittliche Treffgenauigkeit durch den MRT war zwar signifikant höher als jene durch Zufall oder Intuition, betrug jedoch nur 60 bis 70%. Wie hoch die Genauigkeit sein muss, um positive Folgen für Patienten zu haben, ist unklar, da dies von vielen verschiedenen Faktoren abhängt. All diese Faktoren zu untersuchen überschreitet das Ausmaß unserer Studien. Die Unterschiede in Treffgenauigkeit spiegeln möglicherweise Unterschiede in Eigenschaften der Fachausübenden wider. Falls dem so ist, und falls diese Eigenschaften modifizierbar sind, könnten Protokolle entwickelt werden, welche für immerwährende hohe Treffgenauigkeit sorgen würden.

Mit den vorliegenden Studien konnten wir zeigen, dass man wissenschaftliche Prinzipien, zum Beispiel experimentelle Blindheit, Randomisierung, Einsatz von Komparatoren und formelle statistische Analysen, durchaus in der MRT-Forschung anwenden kann. Um weitere Gebiete, zum Beispiel die Notwendigkeit von

Nahrungsergänzungsmitteln oder die Identifizierung von Allergien, Hypersensitivität oder Toxizität, durch den MRT zu erforschen sind zusätzliche Studien erforderlich.

Zukünftige Studien, welche die Brauchbarkeit des MRTs als diagnostisches Instrument erforschen wollen, sollten wissenschaftlich präzise Methoden anwenden: 1) ein klares und spezifisches Forschungsziel, 2) eine klar definierte Zielkondition, 3) explizite, einfach zu interpretierende Ergebnisse, 4) eine angemessene Stichprobe der Zielpopulation (welche objektiv selektiert wurde), 5) eine objektive Vergleichsnorm, 6) eine adäquate Stichprobengröße, und 7) angemessene experimentelle Blindheit.

Schlussendlich lässt sich sagen, dass der klinische Wert des MRTs aufgrund seiner weitverbreiteten Nutzung weiter erforscht werden muss. Die Wirksamkeit verschiedener MRT-Techniken muss in präzisen, randomisierten, kontrollierten Studien untersucht werden. Zum Beispiel könnten zukünftige Wissenschaftler den Wirkungsgrad alternativer stressreduzierender Verfahren, welche den MRT nutzen, wie zum Beispiel HeartSpeak, untersuchen. Diese Techniken werden zum Beispiel bei Depressionen und Panikattacken angewendet und könnten mit traditionellen psychologischen Vorgehen, wie zum Beispiel kognitiver Verhaltenstherapie, verglichen werden.

Schlussfolgerung

Durch die vorliegende Studie konnte gezeigt werden, dass der MRT wahren von unwahren Aussagen genauer unterscheiden kann als sowohl Zufallswahrscheinlichkeit als auch Intuition. Obwohl der Test nicht perfekt scheint, sind die Ergebnisse vielversprechend: Es ist zu hoffen, dass die

vorliegenden Studien zu weiteren Forschungen, welche den klinischen Nutzen des MRTs untersuchen, anregen.