

Die diagnostische Genauigkeit von weit verbreiteten Selbstbeurteilungs-Fragebögen zur Erkennung von Angststörungen - systematische Übersicht und Multi-Test-Metaanalyse (DIAQANDI - 01KG2105)

**Die diagnostische Genauigkeit  
von weit verbreiteten Selbstbeurteilungs-Fragebögen  
zur Erkennung von Angststörungen  
- systematische Übersicht und Multi-Test-Metaanalyse  
(DIAQANDI)**

Sachbericht zum Verwendungsnachweis

Förderkennzeichen: 01KG2105

Zuwendungsempfänger: Technische Universität München

Ausführende Stelle: Institut für Allgemeinmedizin und Versorgungsforschung, Department of  
Clinical Medicine, TUM School of Medicine and Health

Koordinierender Antragsteller: Univ.-Prof. Dr. med. Antonius Schneider

Laufzeit des Vorhabens: 01.07.2021 – 31.12.2023

## **Teil I: Kurzbericht**

### **1. Ursprüngliche Aufgabenstellung sowie wissenschaftlicher und technischer Stand, an den angeknüpft wurde**

Angststörungen sind weltweit die häufigsten psychischen Störungen. Laut epidemiologischen Studien liegt die 12-Monats-Prävalenz zwischen 8% und 21%. Angststörungen bleiben oft unentdeckt, da sich die Symptome mit denen anderer häufig vorkommender Krankheiten überlagern. Selbstauskunftsfragebögen sind eine wichtige Hilfe, um Personen mit Angststörungen zu identifizieren. In solchen Fragebögen werden die Befragten gebeten, das Vorhandensein und den Schweregrad häufiger Angstsymptome zu bewerten. Es gibt eine Fülle unterschiedlicher Bögen, jedoch ist nur eine begrenzte Anzahl dieser Instrumente weit verbreitet und nur wenige wurden ausreichend darauf getestet, ob sie Personen mit Angststörung von Personen ohne Angststörung zuverlässig unterscheiden können. Ziel des Projektes war es deshalb, in systematischen Übersichtsarbeiten die diagnostische Genauigkeit der am weitesten verbreiteten und am häufigsten getesteten Fragebögen im Abgleich mit diagnostischen Standardinterviews zu bestimmen und zu vergleichen. Dies ist erforderlich, um zu beurteilen, a) wie genau vorhandene Instrumente sind und b) ob einige besser als andere sind. Es sollte daher nach allen relevanten Studien zu den wichtigsten Fragebögen gesucht werden, deren Qualität bewertet und die Ergebnisse extrahiert und zusammengefasst werden. Fünf besonders wichtige Fragebögen wurden durch eine systematische Literaturrecherche vorab identifiziert: Die Generalized Anxiety Disorder 7-item Scale 7 (GAD-7) und ihre Kurzform GAD-2, die Hospital Anxiety and Depression Scale - Anxiety Subscale (HADS-A), das Beck Anxiety Inventory (BAI) und das State-Trait Anxiety Inventory (STAI). Eine besondere Herausforderung bei der Interpretation von Fragebögen ist die Wahl des Entscheidungswertes ("Cut-off-Punkt"), ab dem ein Fragebogenergebnis als auffällig gewertet werden soll. Die vier systematischen Übersichtsarbeiten (GAD-7 und GAD-2 wurden in einer Übersicht bearbeitet) sollten daher alle verfügbaren Cut-off-Punkte nutzen, um den diagnostischen Nutzen, aber auch die potenzielle Schäden aufgrund von Unter- und Überdiagnosen in verschiedenen Personengruppen abschätzen zu können. Zudem war ein Vergleich der diagnostischen Genauigkeit mithilfe einer neuen Methode ("Netzwerk-Meta-Analyse") geplant.

### **2. Ablauf des Vorhabens**

Bei der Beantragung war noch nicht sichergestellt, dass die Übersichtsarbeiten im Rahmen der Cochrane Collaboration (<https://www.cochrane.org/de/about-us>) - einem internationalen Netzwerk für besonders hochwertige systematische Reviews - erfolgen sollten. Daher war im Antrag eine flexible Projektlaufzeit von 24 Monaten (ohne Cochrane) plus 6 Monaten (mit Cochrane aufgrund der dort vorgegebenen Prüfungs- und Begutachtungsprozesse) angegeben. Bereits für die Fertigstellung des Protokolls wurden durch die tatsächliche erfolgte Einbindung in Cochrane fünf statt drei Monate benötigt. Die primären Literatursuchen, die Selektionsprozesse, der Aufbau und die Implementierung des Patient-Provider-Boards sowie die Planung der statistischen Analysen erfolgten weitgehend entsprechend der Meilensteinplanung. Dagegen gestaltete sich die Datenextraktion der Studien als aufwändiger als gedacht. Zum einen konnten mehr relevante Studien identifiziert werden als angenommen. Zum anderen waren überraschend viele Studienautoren bereit, uns zusätzliche unveröffentlichte Daten (d.h. Schätzer für diagnostische Genauigkeit zu in der Publikation nicht berichteten Cut-off-Punkten) zur Verfügung zu stellen. Das erhöhte die Qualität des Projektes erheblich, führte aber zu Verzögerungen und zusätzlichem Zeitaufwand. Als größte Herausforderung für Ablauf und Projektmanagement erwies sich jedoch in der letzten Projektphase der intensive und langsame Begutachtungsprozess bei Cochrane. Weitgehend den Planungen entsprechend wurde der erste große

Die diagnostische Genauigkeit von weit verbreiteten Selbstbeurteilungs-Fragebögen zur Erkennung von Angststörungen - systematische Übersicht und Multi-Test-Metaanalyse (DIAQANDI - 01KG2105)

Review zu GAD-7 und GAD-2 im März 2023 zur Veröffentlichung bei Cochrane eingereicht. Eigentlich war geplant, die weiteren Reviews erst einzureichen, wenn die Gutachten zu GAD-7 und GAD-2 vorliegen, um grundsätzliche Kritik am methodischen Vorgehen zu berücksichtigen. Im Oktober 2023 sahen wir uns aber aufgrund der ablaufenden Projektlaufzeit gezwungen, zumindest den zweiten großen Review zur HADS-A einzureichen. Aufgrund einzelner wichtiger Verbesserungsvorschläge des erst im Dezember 2023 eintreffenden (grundsätzlich sehr positiven) Gutachtens zum Review zu GAD-7 und GAD-2, wurden einzelne wichtige Veränderungen für alle vier Reviews vorgenommen. Zum Zeitpunkt der Berichterstellung (15.6.2024) befindet sich der Review zu GAD-7 und GAD-2 in der zweiten Runde des Begutachtungsverfahrens, zum Review zu HADS-A trafen die Gutachten am 10.6.2024 ein. Die Reviews zu BAI und STAI werden im Juli und voraussichtlich im September 2024 eingereicht. Die Durchführung der geplanten Netzwerk-Meta-Analyse kann erst im Herbst 2024 (nach Einarbeitung der letzten Aktualisierungssuchen) erfolgen. Deren Abschluss war jedoch nicht Teil des Meilensteinplanes. Trotz der aufgetretenen Verzögerung ist das Projekt aus Sicht der Projektverantwortlichen sehr erfolgreich gelaufen. Das primäre Projektziel der Durchführung und Einreichung der vier Cochrane-Reviews wird mit einer Verzögerung von acht Monaten erreicht. Die Projektverantwortlichen am Institut für Allgemeinmedizin und Versorgungsforschung der TU München werden Sorge tragen, dass auch die Netzwerk-Metaanalyse fertiggestellt und publiziert werden wird.

### **3. Wesentliche Ergebnisse sowie ggfs. Zusammenarbeit mit anderen Forschungseinrichtungen**

In die vier systematischen Übersichtsarbeiten konnte eine große Zahl diagnostischer Studien eingeschlossen werden, für die zumindestens bei einer der vordefinierten Zielerkrankungen Daten zur Verfügung standen (jegliche Form von Angststörung, generalisierte Angststörung, Panikstörung): 48 Studien mit insgesamt 19,228 Teilnehmern für GAD-7 bzw. GAD-2, 75 Studien mit insgesamt 21,141 Teilnehmern für den HADS-A, 14 Studien mit insgesamt 6232 Teilnehmern für den BAI und 12 Studien mit 2583 Teilnehmern für den STAI. Diese Zahlen können sich aufgrund von Aktualisierungen und eventuellen Änderungen aufgrund ausstehender Begutachtung noch in begrenztem Ausmaß ändern. Ein große Besonderheit der erstellten Übersichtsarbeiten ist, dass für viele Studien diagnostische Daten nicht nur wie sonst üblich diagnostische Daten für einen oder eine kleine Zahl von Cut-off-Werten, sondern für viele bzw. alle möglichen Cuf-off-Werte zur Verfügung standen. Dies erlaubte den breiten Einsatz eines innovativen statistischen Auswertungsmodells (multiple thresholds Modell), das eine umfassende Abschätzung der diagnostischen Genauigkeit für die gesamte Skalenbreite der einzelnen Fragebogens ermöglicht.

Die Ergebnisse unserer Reviews dürften eine erhebliche Rolle bei zukünftigen Diskussionen über die Sinnhaftigkeit von Screeningmaßnahmen für Angsterkrankungen (wie sie 2023 z.B. von der US Preventive Task Force empfohlen wurden) spielen. Grundsätzlich sprechen die Ergebnisse nach formalen Kriterien (Bewertung über die Fläche unter der summary receiver operating characteristic Kurve aus dem multiple thresholds Modell) für eine akzetable bis gute diagnostische Genauigkeit der Fragebögen. Dennoch kann es aber in Abhängigkeit von der tatsächlichen Häufigkeit von Angsterkrankungen in der gescreenten Personengruppe zu einer relevanten Zahl von falsch positiven oder falsch negativen Testergebnissen kommen. Dies macht klar, dass die positiven und negativen Konsequenzen eines Angstscreenings sehr gut miteinander abgewogen werden müssen. Der innovative methodische Ansatz unserer Reviews mit der Verwendung möglichst umfassender diagnostischer Daten könnte Vorbildcharakter für viele zukünftige diagnostische Reviews haben. Eine Durchführung von Reviews im Rahmen der Cochrane Collaboration gewährleistet zwar eine sehr hohe methodische Qualität, ist aber aufgrund der extrem zeitraubenden und zum Teil sehr bürokratischen Prozesse für die Projektplanung problematisch.

## Teil II: Eingehende Darstellung

### 4. Hintergrund und Ziele des Projektes

Angststörungen sind weltweit die häufigsten psychischen Störungen. Laut epidemiologischen Studien liegt die 12-Monats-Prävalenz in der erwachsenen Bevölkerung zwischen 8% und 21% [1,2]. Angststörungen bleiben oft unentdeckt, da sich die Symptome mit denen anderer häufig vorkommender Krankheiten überlagern. Daher haben einzelne Expertengremien – darunter im Jahr 2023 die einflussreiche US Preventive Services Task Force [3] – empfohlen, Screeningprogramme mit Selbstauskunftsfragebögen durchzuführen. In solchen Fragebögen werden die Befragten gebeten, das Vorhandensein und den Schweregrad häufiger Angstsymptome zu bewerten. Es gibt eine Fülle derartiger Fragebögen. Jedoch ist nur eine begrenzte Anzahl dieser Instrumente weit verbreitet und es wurden nur wenige ausreichend in Bezug auf ihre diagnostische Genauigkeit (d.h. ob sie Personen mit Angststörung von Personen ohne Angststörung ausreichend gut unterscheiden) untersucht [4]. Umfassende aktuelle systematische Übersichtsarbeiten (synonym systematische Reviews) zu den Ergebnissen der verfügbaren Studien zur diagnostischen Genauigkeit liegen nicht vor.

Ziel des Projektes war es deshalb, die diagnostische Genauigkeit (mit Sensitivität und Spezifität als deren wichtigsten Maßen) der am weitesten verbreiteten und am häufigsten getesteten Fragebögen, im Abgleich mit diagnostischen Standardinterviews, zu bestimmen und zu vergleichen. Fünf besonders wichtige Fragebögen wurden durch eine systematische Literaturrecherche vorab identifiziert: Die Generalized Anxiety Disorder 7-item Scale 7 (GAD-7) und ihre Kurzform GAD-2, die Hospital Anxiety and Depression Scale - Anxiety Subscale (HADS-A), das Beck Anxiety Inventory (BAI) und das State-Trait Anxiety Inventory (STAI). Beim STAI ist zu beachten, dass es aus zwei getrennt auszuwertenden Subskalen besteht (STAI Trait = STAI-T und STAI State = STAI-S). Alle fünf Instrumente liegen in zahlreichen Sprachen vor, werden klinisch verwendet und wurden in Bezug auf ihre diagnostische Genauigkeit in mindestens zehn Studien untersucht.

Eine besondere Herausforderung bei der Interpretation von Fragebögen ist die Wahl des Entscheidungswertes ("Cut-off-Punkt"), ab dem ein Fragebogenergebnis als auffällig gewertet werden soll. Die vier im Rahmen des vorliegenden Projektes durchgeführten Reviews (GAD-7 und GAD-2 wurden in einem Review bearbeitet, ebenso STAI-S und STAI-T) sollten daher die Daten zur diagnostischen Genauigkeit für alle verfügbaren Cut-off-Punkte optimal nutzen, um untersuchen zu können, wie sich die Aussagekraft in konkreten Anwendungssituationen (in denen positive und negative prädiktive Werte bedeutsamer sind als Sensitivität und Spezifität) in Abhängigkeit vom gewählten Cut-off-Wert verändert. Zudem war ein Vergleich der diagnostischen Genauigkeit in einer Netzwerk-Meta-Analyse geplant. Die beiden letzten Aspekte wurden in systematischen Reviews zu diagnostischen Verfahren bisher nur sehr selten umgesetzt. Damit hatte unser Projekt methodisch eine hohes innovatives Potenzial.

### 5. Einbindung der zu erstellenden Reviews in die Cochrane Collaboration

Die Cochrane Collaboration ([www.cochrane.org](http://www.cochrane.org)), meist verkürzt nur mit Cochrane bezeichnet, ist ein globales, unabhängiges Netzwerk aus Wissenschaftler:innen, Ärzt:innen, Angehörigen der Gesundheitsfachberufe, Patient:innen und weiteren an Gesundheitsfragen interessierten Personen. Primäres Ziel ist die Erstellung, Aktualisierung und Disseminierung hochwertiger, relevanter und aktueller systematischer Übersichtsarbeiten. Für alle Cochrane Reviews müssen vorab Protokolle erstellt werden, die einem intensiven Begutachtungsverfahren unterzogen werden. Bei der Erarbeitung von Protokollen und Reviews müssen zahlreiche Vorgaben eingehalten werden, um eine

Die diagnostische Genauigkeit von weit verbreiteten Selbstbeurteilungs-Fragebögen zur Erkennung von Angststörungen - systematische Übersicht und Multi-Test-Metaanalyse (DIAQANDI - 01KG2105)

hohe Qualität zu garantieren. Außerdem müssen alle Cochrane Reviews mit einer definierten Software (Review Manager) erstellt werden. Die vier im Rahmen des geförderten Projektes durchgeführten systematischen Reviews wurden nach Rücksprache mit dem Projektträger als Cochrane-Reviews durchgeführt. Ein übergreifendes Protokoll wurde 2022 in der Cochrane Database of Systematic Reviews publiziert [5].

Bei der Projektplanung war bereits berücksichtigt worden, dass sich der zeitliche Aufwand für die Durchführung aufgrund der umfangreichen Vorgaben und komplexen Prozesse von Cochrane erhöhen würde (30 Monate statt 24 Monate ohne Cochraneeinbindung). Während der Projektlaufzeit ergaben sich bei Cochrane jedoch tiefgreifende Änderungen, die die Projektabwicklung erheblich beeinflussten und teilweise deutlich erschwerten. So wurde z.B. die vorher für die Betreuung und Koordination der Reviewerstellung zuständige Collaborative Review Group Common Mental Disorder abgewickelt, wodurch eine Rücksprache bei prozessrelevanten Reviewentscheidungen nicht mehr möglich war. Die Begutachtungszeiten beim Protokollreview, insbesondere aber nach Einreichung der fertigen Reviews nahmen ein für das Projektmanagement problematisches Ausmaß an: nach der Einreichung des ersten Reviews (zu GAD-7 und GAD-2) betrug diese acht Monate, für den zweiten Review (zum HADS-A) sieben Monate. Die fehlenden bzw. späten Gutachten hatten zur Folge, dass relevante Gutachterkritik für die Fertigstellung der Folgereviews sehr spät (für BAI und STAI) bzw. gar nicht (für HADS-A) zu Verfügung stand. Die Gutachterkommentare in Cochrane Reviews sind erfahrungsgemäß sehr umfangreich, methodisch hochwertig und anspruchsvoll in der Bearbeitung. In der Tat erfolgte aufgrund der Rückmeldungen auf die Ersteinreichung des Reviews zu GAD-7 und GAD-2 nochmals eine Schärfung der Ein- und Ausschlusskriterien mit erheblichen Konsequenzen (siehe Abschnitt 4). Außerdem wurde die Reviewerstellungsoftware während der Prozesslaufzeit in ein reines online-Tool umgewandelt, was unter anderem zur Folge hatte, dass die vorher übliche automatische Registrierung von Cochrane Protokollen in dem Review-Protokoll-Portal PROSPERO nicht mehr möglich war. Die Einbindung in Cochrane in das vorliegende Projekt muss daher aus heutiger Sicht sehr ambivalent bewertet werden: für die methodische Qualität sehr positiv, für Projektplanung und -management sehr belastend und verzögernd.

## **6. Identifikation, Selektion, Datenextraktion und methodische Bewertung der Primärstudien**

Durch die - für die vier Reviews getrennt durchgeführten - Literatursuchen wurden insgesamt 26.727 Quellen identifiziert, von denen allerdings 8.950 Duplikate waren (siehe Abbildung 1 für die Flussdiagramme der vier Reviews). In das durch zwei unabhängig voneinander arbeitenden Reviewern durchgeführte Titel- und Abstractscreening wurden somit 17.777 Quellen einbezogen, von denen 16.842 als eindeutig irrelevant verworfen wurden. Für die verbleibenden 935 Quellen wurden die Volltexte besorgt und diese gegen die Selektionskriterien geprüft. Diese waren für 748 nicht erfüllt. Letztlich wurden in den GAD-7- und GAD-2-Review 48 Studien (berichtet in 63 relevanten Publikationen) eingeschlossen, in den HADS-A-Review 75 Studien (100 Publikationen), in den BAI-Review 14 Studien (18 Publikationen) und den STAI-Review 13 Studien (13 Publikationen). Da einzelne Studien mehr als einen Bogen untersuchten, ist die Gesamtzahl der einbezogenen Studien nicht die Summe der Studien pro Review (149), sondern etwas geringer (133). Alle hier berichteten Zahlen sind zum Zeitpunkt der Berichterstellung (25. Juni 2024) aktuell; sie können in zukünftigen Publikationen allerdings aufgrund von Änderungen aufgrund der Begutachtungsverfahren bei Cochrane und notwendigen Update-Suchen abweichen.

## Die diagnostische Genauigkeit von weit verbreiteten Selbstbeurteilungs-Fragebögen zur Erkennung von Angststörungen - systematische Übersicht und Multi-Test-Metaanalyse (DIAQANDI - 01KG2105)

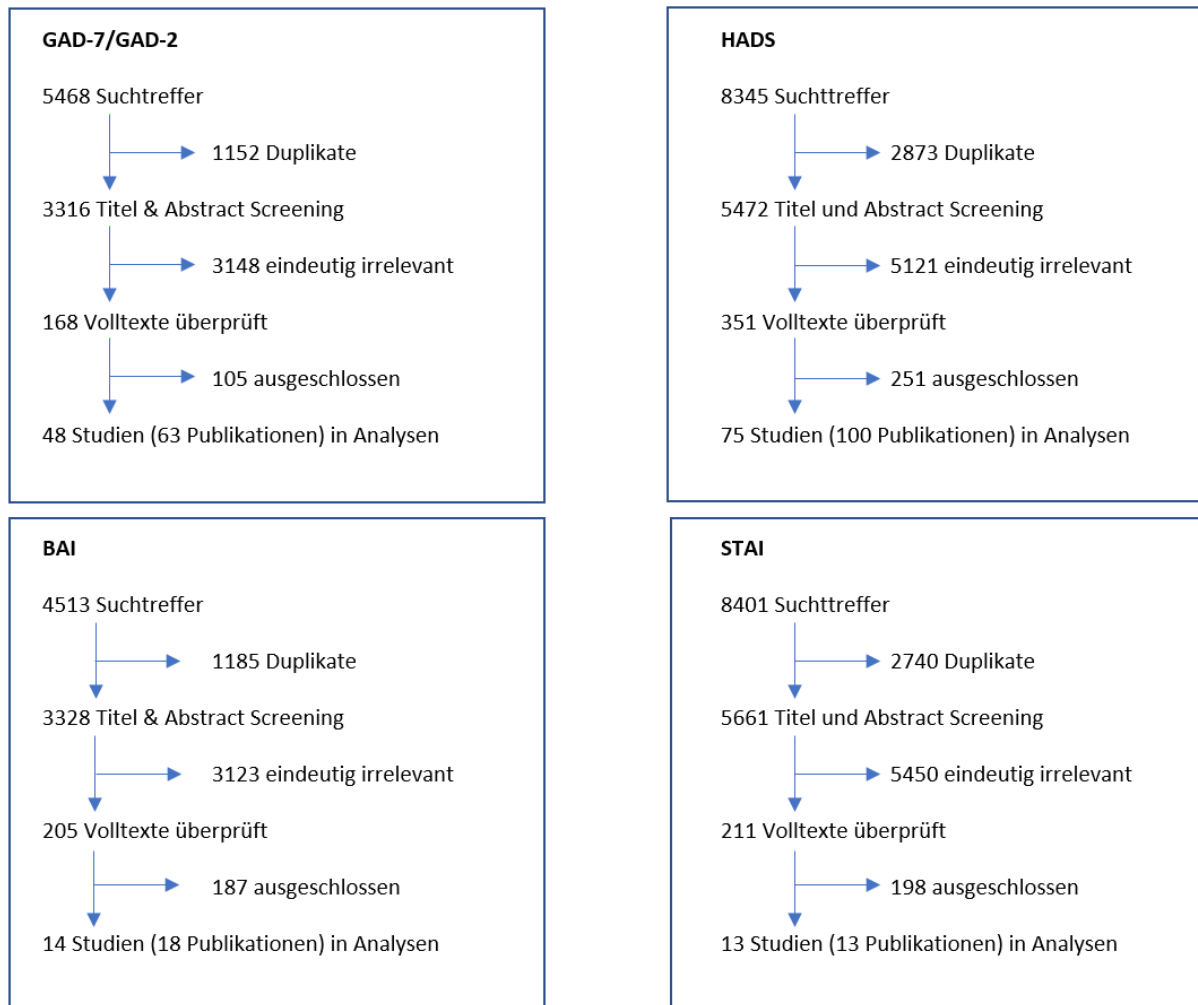


Abbildung 1 - Flussdiagramme der vier systematischen Reviews (Stand 25. Juni 2024)

Alle eingeschlossenen Studien wurden einer Datenextraktion und einer methodischen Bewertung durch mindestens zwei unabhängig voneinander arbeitende Reviewer unterzogen. Bei der Datenextraktion wurden Informationen zu Patient:innen, Studiendesign, Index Tests, Referenzstandard und Auswertung in einer Exceldatei dokumentiert, vor allem aber die Studienergebnisse zusammengefasst. Bei den Studienergebnissen wurden insbesondere Angaben zu Sensitivität und Spezifität zu allen berichteten Cut-off-Punkten dokumentiert; letztlich wurden für die Analysen Vierfeldertafeln generiert. Die methodische Qualität (genauer das Bias-Risiko und die „Applicability“) wurden mit dem von Cochrane empfohlenen QUADAS-II-Tool bewertet.

Da die Berichterstattung bei vielen Studien Fragen offen ließ, mussten für fast alle Studien die Autor:innen kontaktiert und um Zusatzinformationen gebeten werden. Dies betraf häufig methodische Details (z.B. zur Zeitspanne zwischen Index Text und Referenzstandard oder zur Qualifikation der Interviewer für den Referenzstandard). Vor allem aber wurde große Mühe darauf verwendet, von den Autor:innen Daten zu Sensitivität und Spezifität von nicht berichteten Cut-off-Punkten zu erhalten. Von 120 angeschriebenen Autor:innen erhielten wir in 72 Fällen mindestens eine Antwort. Bei Nicht-Antwort erfolgte mindestens ein Erinnerungsschreiben. Umfang und Qualität der Antworten variierten jedoch zwischen den Autor:innen erheblich. Erfreulicherweise waren fast zwei Drittel der antwortenden Autor:innen bereit, umfangreiche Zusatzinformationen zu liefern. Insbesondere wenn die Autor:innen bereit waren, zusätzliche Ergebnisdaten zur Verfügung zu stellen (und hierfür

Die diagnostische Genauigkeit von weit verbreiteten Selbstbeurteilungs-Fragebögen zur Erkennung von Angststörungen - systematische Übersicht und Multi-Test-Metaanalyse (DIAQANDI - 01KG2105)

eventuell ihre Daten nochmals analysieren mussten), konnte dies zu einer aufwändigen und langwierigen E-Mail-Kommunikation führen. In Bezug auf den großen Arbeitsaufwand ist außerdem zu berücksichtigen, dass festgelegt war, mindestens für zwei besonders häufig berichtete „Zielerkrankungen“ (d.h. die durch den Referenzstandard klassifizierten und berichteten Diagnosen) Daten zu extrahieren. In den vorhandenen Studien waren dies typischerweise die Sammelkategorie „any anxiety disorder“ (jegliche Art von Angststörung) und die spezifische Störung „generalized anxiety disorder“ (generalisierte Angststörung). In zwei Reviews (HADS-A und BAI) wurde zusätzlich auch die Zielerkrankung „panic disorder“ (Panikstörung) bearbeitet, da hierfür ausreichend viele Studien Ergebnisse berichteten.

## 7. Exemplarische Einblicke in die Analysen und deren Ergebnisse

Da im vorliegenden Projekt insgesamt fünf Instrumente bzw. Fragebögen mit sechs Skalen (beim STAI die beiden Subskalen STAI-T und STAI-S) für zwei (GAD-7, GAD-2, STAI-T und STAI-S) oder drei Zielerkrankungen (HADS-A und BAI) untersucht wurden, erfolgten 14 primäre Analysen zu Sensitivität und Spezifität.

Üblicherweise werden in systematischen Reviews von diagnostischen Genauigkeitsstudien Metaanalysen nur für einen einzigen Cut-off-Punkt oder eine begrenzte Zahl von Cut-off-Punkten durchgeführt. Eine besondere Stärke des vorliegenden Projektes ist, dass Daten zur Sensitivität und Spezifität zu so vielen Cut-off-Punkten wie möglich gesammelt und ausgewertet wurden (siehe Tabelle 1). Dies wurde dadurch möglich, dass neben dem gängigen bivariaten Modell, bei dem pro Studie und Zielerkrankung immer nur eine Vierfeldertafel in eine Analyse eingehen kann, das multiple thresholds Modell [6] verwendet wurde, bei dem alle verfügbaren Cut-off-Werte eingehen. Die prinzipielle Anwendbarkeit des multiple thresholds Modells wurde in einem früheren systematischen Review zur diagnostischen Genauigkeit von fraktioniertem exhalierendem Stickstoffmonoxid (FeNO) bei Verdacht auf Asthma bronchiale gezeigt [7].

Insgesamt gingen 3351 diagnostische Vierfeldertafeln in die Analysen ein.

Tabelle 1

Anzahl der Studien / Vierfeldertafeln pro Fragebogen und Zielerkrankung

Instrument (Skalenbreite)	Jegliche Art von Angststörung	Generalisierte Angststörung	Panikstörung	Vierfeldertafeln gesamt
GAD-7 (0-21)	22 / 308	35 / 447	-	755
GAD-2 (0-6)	19 / 94	24 / 117	-	211
HADS-A (0-21)	61 / 627	37 / 465	10 / 138	1230
BAI (0-63)	10 / 302	8 / 201	4 / 113	616
STAI-S (20-80)	10 / 160	3 / 123	-	283
STAI-T (20-80)	8 / 182	2 / 74	-	256

Laut Protokoll [5] sollten nach Möglichkeit die primären Analysen mit Hilfe des bivariaten Modells für den für die jeweiligen Skalen empfohlenen Cut-off-Punkt (oder falls für diesen Daten keine Ergebnisse berichtet waren, für den nächstgelegenen berichteten Cut-off-Punkt, wenn dieser in einem vordefinierten Bereich lag) erfolgen. Dies konnte für GAD-7, GAD-2 und HADS-A für alle jeweils

Die diagnostische Genauigkeit von weit verbreiteten Selbstbeurteilungs-Fragebögen zur Erkennung von Angststörungen - systematische Übersicht und Multi-Test-Metaanalyse (DIAQANDI - 01KG2105)

untersuchten Zielerkrankungen (also in insgesamt 7 von 14 primären Analysen) umgesetzt werden. Abbildung 2 zeigt als Beispiel die Ergebnisse der bivariaten Analyse zur Sensitivität und Spezifität des GAD-7 zur Erkennung bzw. dem Ausschluss einer generalisierten Angststörung. Über alle Studien hinweg lag der Gesamtschätzer für die Sensitivität bei 0,64 (95%-Konfidenzintervall 0,56-0,72), für die Spezifität bei 0,91 (0,87-0,93). Die Schätzer in den einzelnen Studien differierten enorm: zwischen 0,09 und 0,91 für die Sensitivität und zwischen 0,60 und 0,99 für die Spezifität.

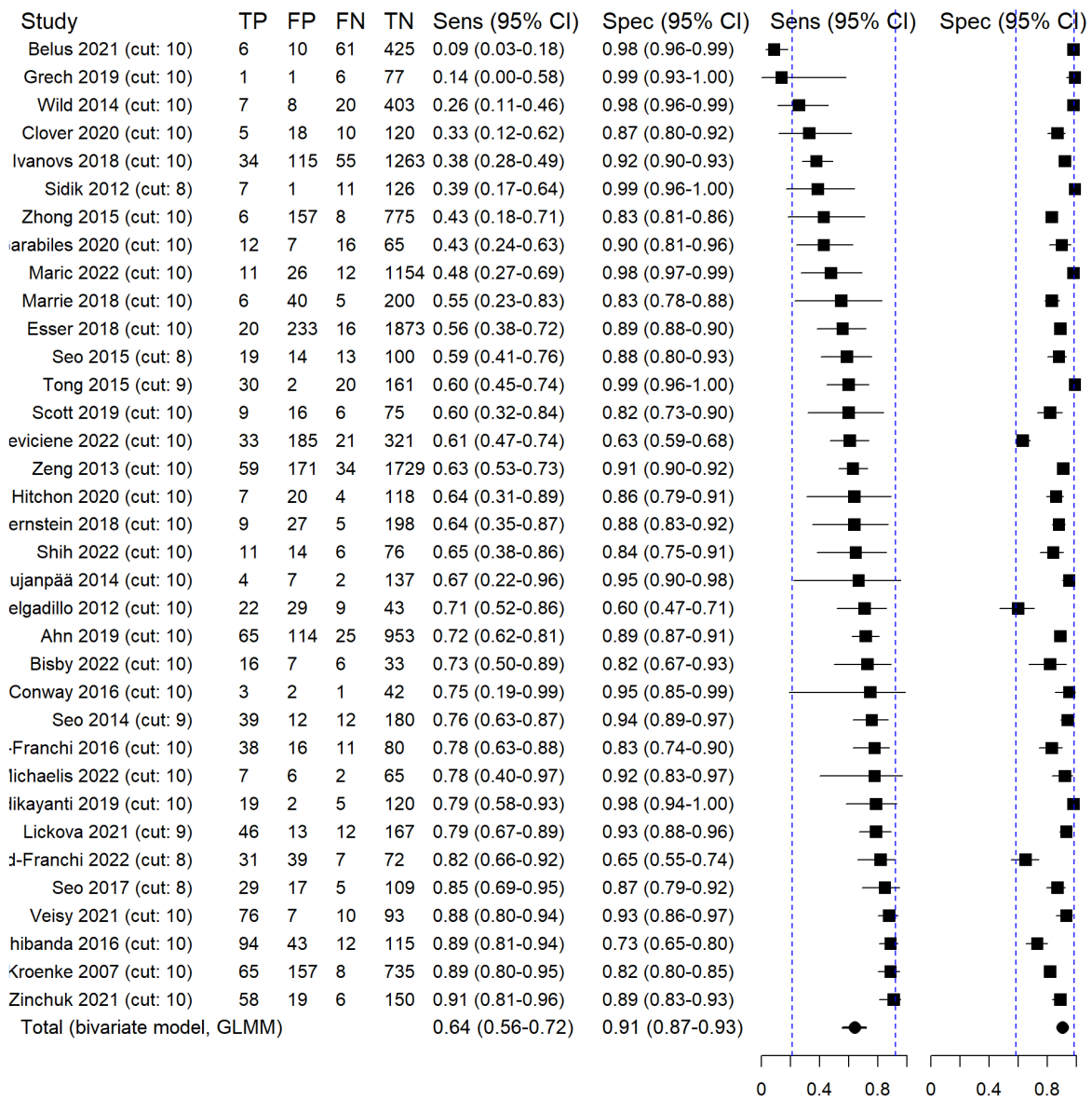


Abbildung 2 - Forest Plot und bivariate Analyse zur Sensitivität und Spezifität des GAD-7 am empfohlenen Cut-off-Punkt  $\geq 10$  oder dem nächstgelegenen Cut-off-Punkt zur Erkennung von generalisierter Angststörung (Stand 25. Juni 2024)

TP = richtig Positive, FP = falsch Positive, FN = falsch Negative, TN = richtig Positive, 95% CI = 95% Konfidenzintervall

Die diagnostische Genauigkeit von weit verbreiteten Selbstbeurteilungs-Fragebögen zur Erkennung von Angststörungen - systematische Übersicht und Multi-Test-Metaanalyse (DIAQANDI - 01KG2105)

Ähnlich große Unterschiede zwischen den Schätzern der einzelnen Studien wurden auch in der Mehrzahl der anderen Analysen beobachtet. Diese Heterogenität wirft die Frage auf, inwiefern die Gesamtschätzer aus den Metaanalysen klinisch sinnvoll interpretierbar sind und wie die Heterogenität erklärt werden kann. Erhebliche Heterogenität ist in diagnostischen Metaanalysen häufig und war daher bis zu einem gewissen Grad erwartet [8]. Im Protokoll waren vorab vier Faktoren definiert worden, die als mögliche Ursachen für die Heterogenität untersucht werden sollten: 1) Das Setting (unterteilt in Studien a) außerhalb eines medizinischen Settings, b) bei Patient:innen unabhängig von der Diagnose oder c) bei Patient:innen mit einer bestimmten Grunderkrankung wie z.B. Epilepsie); 2) die Prävalenz der Zielerkrankung; 3) die Art des als Referenzstandard verwendeten diagnostischen Interviews; und 4) das Biasrisiko.

Insbesondere das Setting hatte in vielen Analysen einen statistisch signifikanten Einfluss auf die beobachtete Heterogenität. Dies führte nach Gutachterkritik in der ersten Runde des Peer Reviews für den GAD-7- und GAD-2-Review zu der Entscheidung, in allen Reviews die besonders auffällige Gruppe der Studien auszuschließen, die in einem psychiatrischen Setting rekrutiert wurden oder bei denen bereits nur Personen mit konkreten Hinweisen auf eine mögliche Angsterkrankung bestanden. Für die Prävalenz und die Art des Referenzstandards ergaben sich vereinzelt Hinweise für einen Einfluss auf die Heterogenität. Für das Biasrisiko war kein Hinweis erkennbar. Subgruppenanalysen wurden für alle vier Faktoren durchgeführt. Hierbei war die Zahl der Studien pro Subgruppe häufig gering, sodass eine zuverlässige Interpretation zum Teil nicht möglich war. Innerhalb ausreichend großer Subgruppen war die Heterogenität meist weiterhin erheblich. Zusammenfassend ist festzustellen, dass die beobachtete Heterogenität durch die vordefinierten Faktoren nur unzureichend erklärt werden kann. Die eingeschlossenen Studien kommen aus einer Vielzahl von Ländern mit unterschiedlichen Sprachen und die Studienpopulationen waren auch innerhalb formal ähnlicher Settings häufig sehr unterschiedlich. Der Einfluss der kaum sinnvoll gruppierbaren sprachlichen und soziokulturellen Faktoren ist aber trotz der insgesamt großen Studienzahl im durchgeführten Projekt nicht zuverlässig empirisch prüfbar.

In den Reviews zu BAI und STAI kam es aufgrund der kleinen Zahl von Studien zu Problemen bezüglich der Parameterschätzung des bivariaten Modells. Wie auch im Cochrane Handbook beschrieben [8], können in solchen Fällen sogenannte „boundary fits“ auftreten, d.h. Schätzwerte, die allgemein als nicht verlässlich angesehen werden, da sie auf dem Rand des möglichen Wertebereichs liegen. Als mögliche Lösungen werden Anpassungen der Modellspezifikation und alternative Modellierungsarten diskutiert. Daher wurde in diesen Reviews das multiple thresholds Modell, das in den Review zu GAD-7, GAD-2 und HADS-A nur für sekundäre Analysen eingesetzt wurde, für die primären Analysen verwendet. Aus Sicht des Berichtsaufstellers ist das multiple thresholds Modell jedoch für alle Reviews die sinnvollere Analysemethode, die konsistentere Parameterschätzer liefert. Abbildung 3 gibt nicht nur einen umfassenden Überblick über die vorhandenen Daten, sondern zeigt auch die zusammenfassenden Schätzer und Konfidenzintervalle an allen Cut-off-Punkten. Ein weiterer Vorteil des multiple thresholds Modells ist die Möglichkeit mit der area under the curve (AUC, Fläche unter der sROC Kurve) ein Gesamtmaß der diagnostischen Genauigkeit zu berechnen. Es gibt keine allgemein anerkannte Einteilung für die Interpretation von AUCs; die in unseren bisherigen Analysen beobachteten Werte zwischen 0,72 und 0,86 werden bei ROC-Analysen üblicherweise als akzeptabel (0,70-79) bis gut (0,80-0,89) angesehen.

Um die Transparenz und Nachvollziehbarkeit zu gewährleisten, werden bei den bei Cochrane eingereichten Reviews alle Vierfeldertafeln sowie die Ergebnisse der Bewertung der methodischen Qualität und der Subgruppenanalysen für alle Studien in den Anhängen präsentiert.

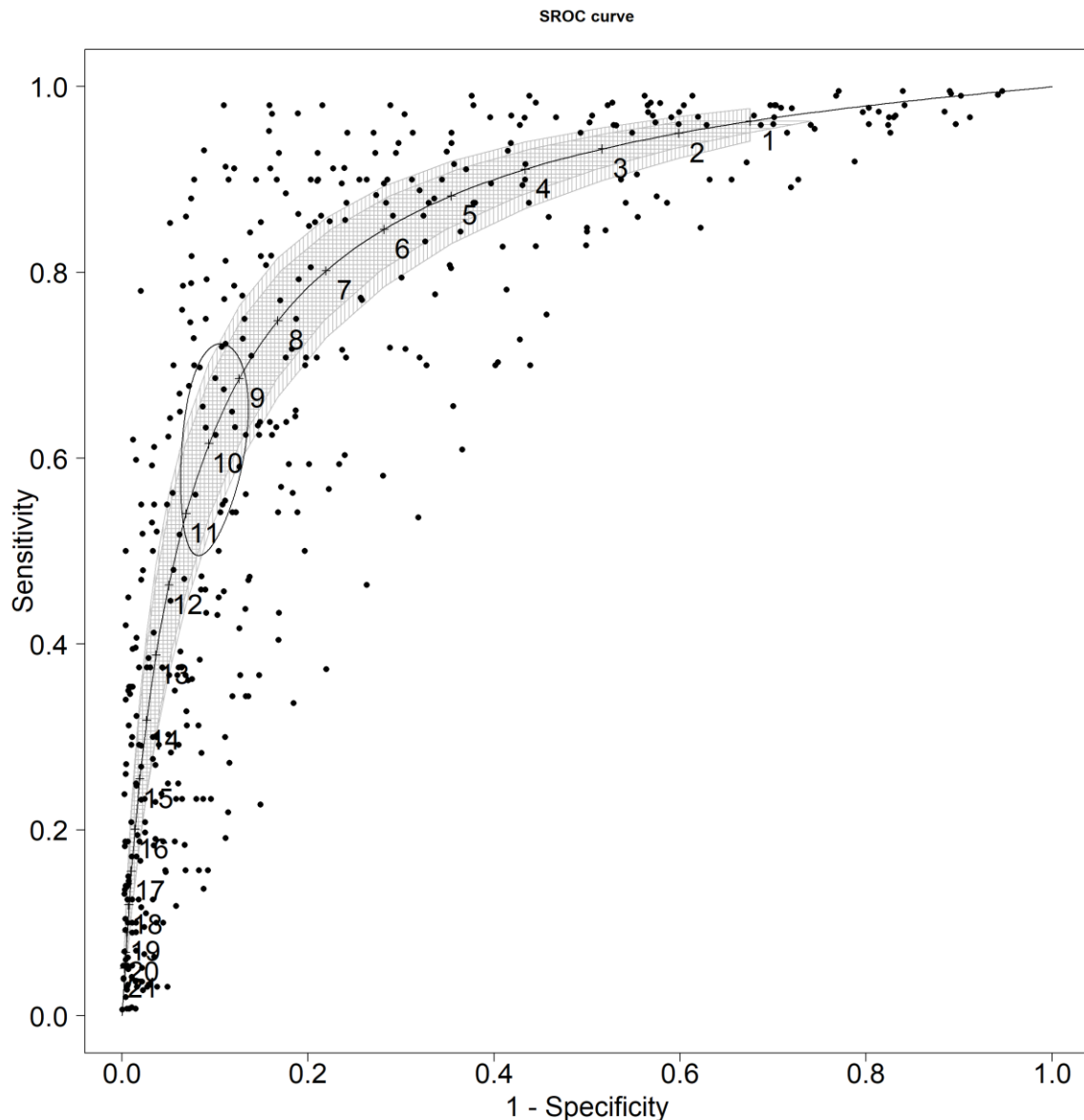


Abbildung 3 - Beispiel multiple thresholds sROC-Kurve (summary receiver operating characteristic): Sensitivität und 1-Spezifität des GAD-7 zur Erkennung von generalisierter Angststörung

Die Punkte stellen alle verfügbaren Sensitivitäts-/1-Spezifitäts-Wertpaare der einzelnen Studien an allen verfügbaren Cut-off-Punkten dar. Die schwarze Linie ist die sROC-Kurve, die Kreuze auf der Kurve markieren die summary Sensitivitäts-/1-Spezifitäts-Wertpaare für die jeweils angegebenen Cut-off Punkte. In den schraffierten Bereichen stellen die waagrechten Linien die 95%-Konfidenzintervalle für die jeweiligen Schätzer für die Sensitivität dar, die senkrechten Linien für die Spezifität.

Um die Interpretation der Ergebnisse für mögliche Screeningsituationen zu erleichtern, wurden systematisch Szenarien erstellt. Abbildung 4 zeigt exemplarisch ein Szenario, in dem die Ergebnisse der Analysen zum GAD-7 zur Erkennung einer generalisierten Angststörung in einer Population von 1000 Personen angewendet werden, in der die Prävalenz dieser Erkrankung 12 % beträgt (dies entspricht der medianen Prävalenz in den eingeschlossenen Studien). Bei der beobachteten Sensitivität und Spezifität von 64 % bzw. 91 %, würde bei 16 % (genau 156 Personen) der Gescreenten ein „positives“ Ergebnis erwartet werden. Diese positiv gescreenten Personen müssten dann einer angemessenen klinischen Diagnostik unterzogen werden. Nur bei der Hälfte (genau 77) der positiv gescreenten Personen, würde jedoch tatsächlich eine generalisierte Angststörung vorliegen, bei der anderen Hälfte

Die diagnostische Genauigkeit von weit verbreiteten Selbstbeurteilungs-Fragebögen zur Erkennung von Angststörungen - systematische Übersicht und Multi-Test-Metaanalyse (DIAQANDI - 01KG2105)

dagegen nicht. Bei den 84 % (genau 844) negativ gescreeenten Personen würde keine weitere Diagnostik erfolgen; bei 43 dieser Personen würde jedoch eine tatsächlich vorliegende generalisierte Angststörung übersehen werden. Dies bedeutet gleichzeitig, dass 36 % (43 von 120) der tatsächlich vorliegenden Angststörungen nicht erkannt werden würden.

Derartige Szenarien ermöglichen eine Abschätzung der möglichen positiven und negativen Konsequenzen von Screeningmaßnahmen. Aufgrund der Heterogenität der vorliegenden Studien und dem Einfluss der Prävalenz der Zielerkrankungen auf die Ergebnisse der Szenarien müssen diese jedoch mit angemessener Zurückhaltung interpretiert werden.

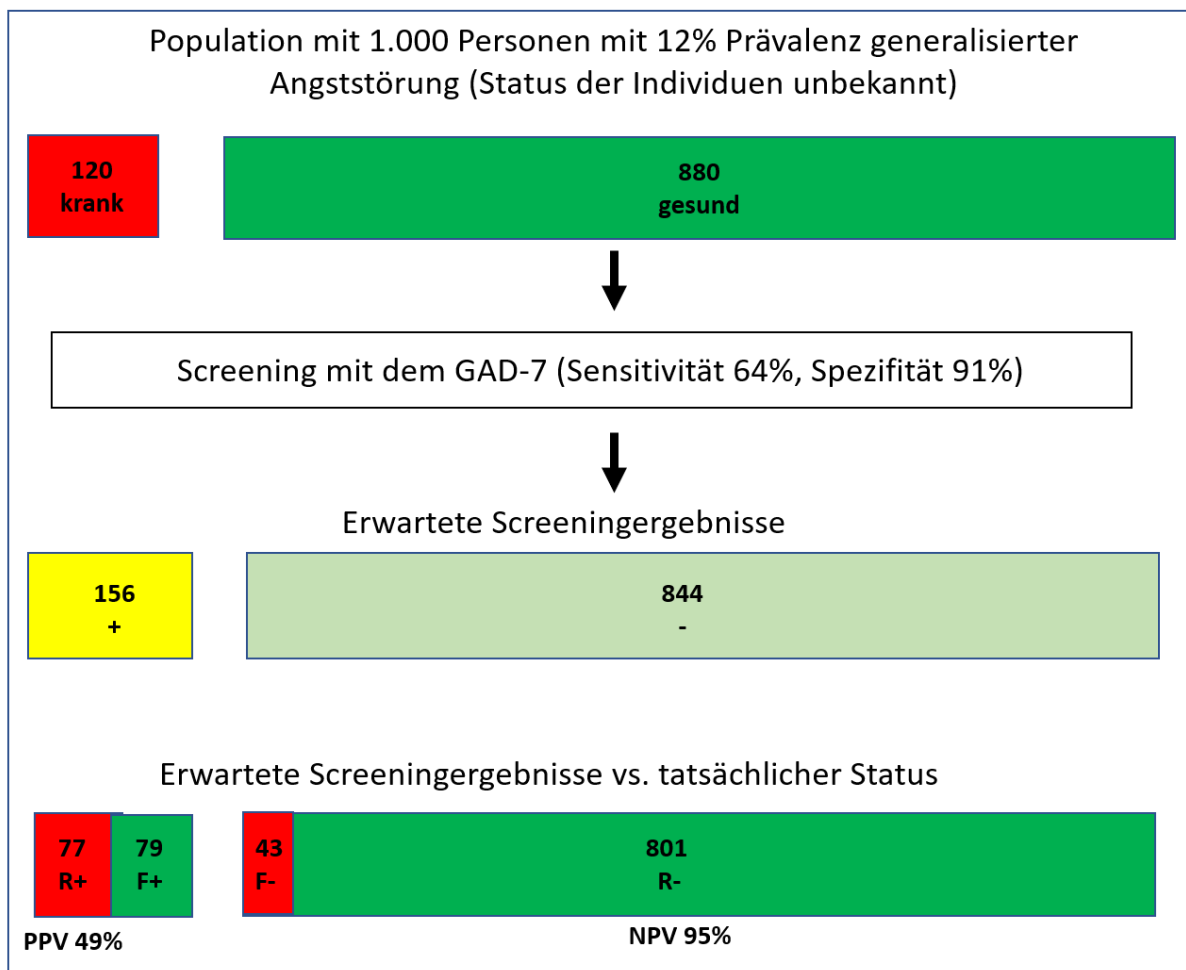


Abbildung 4 - Szenario für ein Angstscreening mit dem GAD-7 (Cut-off-Punkt  $\geq 10$  oder nächstgelegener Cut-off-Punkt) zur Erkennung von Personen mit einer generalisierten Angststörung in einer Population mit einer Krankheitsprävalenz von 12% (mediane Prävalenz in den analysierten Studien) auf Basis der in der bivariaten Metaanalyse ermittelten Sensitivität und Spezifität.

R+ = richtig positives Screeningergebnis, F+ = falsch positiv, F- = falsch negativ, R- = richtig negativ; PPV = positiver prädiktiver Wert, NPV = negativer prädiktiver Wert

## 8. Ausblick Netzwerk-Metaanalyse

Durch den Einbezug von Cochrane bei den vier durchgeführten systematischen Reviews war es nicht mehr möglich, die geplante Netzwerk-Metaanalyse innerhalb der Projektlaufzeit fertigzustellen. Um

Die diagnostische Genauigkeit von weit verbreiteten Selbstbeurteilungs-Fragebögen zur Erkennung von Angststörungen - systematische Übersicht und Multi-Test-Metaanalyse (DIAQANDI - 01KG2105)

eine methodisch hochwertige und aktuelle Netzwerk-Metaanalyse vorlegen zu können, erschien es den Projektverantwortlichen zunächst notwendig, Gutachterrückmeldungen zu mindestens einem der nach dem gleichen Protokoll durchgeführten Cochrane-Reviews abzuwarten, um relevante Anpassungen vornehmen zu können. Durch das langwierige Peer-review-Verfahren ergab bzw. ergibt sich weiterhin das Problem, dass die vor der Ersteinreichung aktualisierten Suchen teilweise bereits wieder veraltet sind und ein weiteres Mal aktualisiert werden müssen. Diese Aktualisierungssuchen gewährleisten, dass bei der Publikation der Cochranereviews und bei der Durchführung des Netzwerk-Metaanalyse eine möglichst aktuelle Studienlage berücksichtigt wird.

Die Datenbanken der vier Reviews sind so konzipiert, dass sie nach den anstehenden letzten Aktualisierungen und Extraktionen zusammengeführt werden können, um die noch ausstehende Netzwerk-Metaanalyse durchzuführen. Eine Probeversion der zusammengeführten Datenbank auf Basis des aktuellen Standes wird aktuell erstellt, um die statistische Auswertung vorzubereiten.

Die für den Herbst 2024 geplante Durchführung der Netzwerk-Metaanalyse wird einen weitgehenden formalen Vergleich der diagnostischen Genauigkeit der untersuchten Instrumente ermöglichen. Allerdings gehen die Projektverantwortlichen davon aus, dass aufgrund der Diversität und Heterogenität der eingeschlossenen Studien die Ergebnisse zurückhaltend interpretiert werden müssen.

#### **9. Während der Durchführung des Vorhabens dem Zuwendungsempfänger bekannt gewordener Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen**

Während der Durchführung wurden keine vergleichbaren systematischen Übersichtsarbeiten von anderen Arbeitsgruppen zu den vom Zuwendungsempfänger bearbeiteten Fragestellungen publiziert. Insbesondere die viel beachteten und diskutierten Empfehlungen der US Preventive Services Task Force [3] für ein Angsst Screening bei Erwachsenen unter 65 Jahren im Jahre 2023 zeigen jedoch die Aktualität des Themas. Als Entscheidungsgrundlage für diese Empfehlung wurde zwar ein „Evidence Report and Systematic Review“ [4] vorgelegt. Die diagnostische Genauigkeit von Screeningfragebögen wurde in diesem Report jedoch nur eingeschränkt untersucht. Aufgrund von sehr starken Settingbegrenzungen wurde nur eine sehr kleine Zahl von diagnostischen Genauigkeitsstudien zu GAD-7, GAD-2 und einzelnen anderen Bögen analysiert. In die durchgeführten Metaanalysen waren maximal drei Studien pro Fragebogen eingegangen. Der HADS-A als das insgesamt am häufigsten untersuchte Instrument wurde gar nicht berücksichtigt.

Im Zuwendungszeitraum neu publizierte diagnostische Genauigkeitsstudien zu den untersuchten Fragebögen wurden auf der Basis von Aktualisierungssuchen berücksichtigt und werden auch derzeit (nach Ablauf der Förderung) so weit möglich noch in die Überarbeitungen der Cochrane Reviews im Rahmen des Begutachtungsprozesses eingearbeitet.

Die diagnostische Genauigkeit von weit verbreiteten Selbstbeurteilungs-Fragebögen zur Erkennung von Angststörungen - systematische Übersicht und Multi-Test-Metaanalyse (DIAQANDI - 01KG2105)

## 10. Ergebnisverwertung/Voraussichtlicher Nutzen

Die unmittelbare Ergebnisverwertung des Projektes erfolgt primär über Publikationen. Bis zum Zeitpunkt der Berichtserstellung wurden neben der Protokollpublikation [5] vorläufige Ergebnisse in drei Präsentationen auf Kongressen vorgestellt [9-11]. Außerdem wurde auf Basis der Erfahrungen des Projekts ein Kongress-Workshop für diagnostische Reviews durchgeführt [12]. Für die vier Reviews ist die Publikation in der Cochrane Database of Systematic Reviews vorgesehen. Anders als bei der Antragserstellung prognostiziert, konnten auch die Ergebnisse des multiple thresholds Modells in die Cochrane Reviews integriert werden. Bei zwei der vier Reviews ist die Einreichung bereits erfolgt, für die anderen zwei Reviews steht die Einreichung unmittelbar (BAI) bzw. in wenigen Monaten (STAI) bevor. Leider ist das Begutachtungsverfahren langsam, langwierig und aufwändig (siehe Abschnitt 2). Für die geplante Netzwerk-Metaanalyse ist die Einreichung bei einem hochrangigen internationalen Journal vorgesehen. Eine Einreichung erscheint aber nicht vor Dezember 2024 realistisch.

Tabelle 2

Aus dem Projekt hervorgehende Hauptpublikationen

	Aktueller Stand (25. Juni 2024)
Protokoll Cochrane Reviews	Publiziert 9/2022 [5]
Cochrane Review GAD-7 und GAD-2	Ersteinreichung 3/2023 - revidierte Version im Peer Review
Cochrane Review HADS-A	Ersteinreichung 10/2023 - im Peer Review
Cochrane Review BAI	Ersteinreichung 7/2024
Cochrane Review STAI	Ersteinreichung geplant 9/2024
Netzwerk-Metaanalyse	Durchführung und Manuskripterstellung nach Einarbeitung der letzten Aktualisierungen voraussichtlich 10-12/2024

Darüber hinaus ist die Erstellung von mindestens zwei weiteren methodisch orientierten Manuskripten geplant, die auf den erhobenen Daten basieren.

Es ist davon auszugehen, dass die erstellten Cochrane Reviews und die geplante Netzwerk-Metaanalyse bei der Erarbeitung von Empfehlungen z.B. in Leitlinien und allgemein bei Diskussionen zu Screening auf Angststörungen einen wichtigen Einfluss haben werden. Wichtig erscheint hier insbesondere, dass die im Projekt erarbeiteten Reviews deutlich machen, dass in den meisten Settings beim Screening mit einer hohen Zahl von Fehlklassifikationen (falsch positiven und/oder falsch negativen Befunden) zu rechnen ist.

Die Projektverantwortlichen gehen außerdem davon aus, dass die erarbeiteten Reviews auch methodisch wichtige Impulse für zukünftige systematische Übersichtsarbeiten zu diagnostischen Genauigkeitsstudien geben werden. Die Sammlung von Ergebnissen über die gesamte Skalenbreite und deren effiziente Analyse durch das multiple thresholds Modell haben sich als gut durchführbar und in hohem Maße informativ erwiesen.

## 11. Wichtigste Positionen des zahlenmäßigen Nachweises

Die Position 0812 Personalausgaben für Beschäftigte mit Eingruppierung E12-E15 war mit 96% der mit Abstand größte Anteil der Projektausgaben. Die im Gesamtfinanzierungsplan vorgesehene Summe wurde um 2% überschritten, was aber durch Einsparungen in den anderen Positionen ausgeglichen

Die diagnostische Genauigkeit von weit verbreiteten Selbstbeurteilungs-Fragebögen zur Erkennung von Angststörungen - systematische Übersicht und Multi-Test-Metaanalyse (DIAQANDI - 01KG2105)

werden konnte. Auch gab es innerhalb der Positionen geringfügige Abweichungen von der Planung. Durch die unerwartet große Zahl einzuschließender Studien und die aufwändige Extraktion und Autorenkommunikation entstand hierfür eine ungeplante Mehrarbeit. Dies konnte jedoch durch einen etwas reduzierten Aufwand bei der Statistik ausgeglichen werden, da die statistischen Berechnungen schneller umzusetzen waren als ursprünglich erwartet. Diese Änderungen wurden dem Projektträger am 27.07.2023 mitgeteilt und am 01.08.2023 genehmigt.

Die Ausgaben in Position 0843 (Vergabe von Aufträgen) und Position 843 (sonstige allgemeine Verwaltungsausgaben) konnten geringer als geplant gehalten werden. In der Position 0846 (Dienstreisen) konnten die Mittel nicht verausgabt werden.

## **12. Notwendigkeit und Angemessenheit der geleisteten Projektarbeiten**

Alle durchgeführten Arbeiten waren für das Erreichen der Vorhabensziele notwendig. Die geleisteten Arbeiten entsprechen hohen Standards für die Erstellung systematischer Übersichtsarbeiten. Die Angemessenheit wurde auch durch die Einbindung der vier Reviews in die Cochrane Collaboration gewährleistet.

## **13. Zitierte Literatur**

1. Kessler RC, Keller MB, Wittchen HU. The epidemiology of generalized anxiety disorder. *Psychiatric Clinics of North America* 2001;24(1):19-39.
2. Remes O, Brayne C, Van Der Linde R, Lafortune L. A systematic review of reviews on the prevalence of anxiety disorders in adult populations. *Brain and Behavior* 2016;6(7):e00497.
3. US Preventive Services Task Force. Screening for Anxiety Disorders in Adults: US Preventive Services Task Force Recommendation Statement. *JAMA* 2023/06/27;329(24):2163-2170.
4. O'Connor EA, Henninger ML, Perdue LA, Coppola EL, Thomas RG, Gaynes BN. Anxiety Screening: Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA* 2023/06/27;329(24):2171-2184.
5. Linde K, Olm M, Teusen C, Aktürk Z, von Schrottenberg V, Hapfelmeier A, et al. The diagnostic accuracy of widely used self-report questionnaires for detecting anxiety disorders in adults. *Cochrane Database of Systematic Reviews* 2022, Issue 9. Art. No: CD015292.
6. Steinhauser S, Schumacher M, Rücker G. Modelling multiple thresholds in meta-analysis of diagnostic test accuracy studies. *BMC Med Res Methodol.* 2016 Aug 12;16(1):97.
7. Schneider A, Linde K, Reitsma JB, Steinhauser S, Rücker G. A novel statistical model for analyzing data of a systematic review generates optimal cutoff values for fractional exhaled nitric oxide for asthma diagnosis. *J Clin Epidemiol.* 2017;92:69-78.
8. Takwoingi Y, Dendukuri N, Schiller I, Rücker G, Jones HE, Partlett C, Macaskill P. Chapter 10: Undertaking meta-analysis. In: Deeks JJ, Bossuyt PM, Leeflang MM, Takwoingi Y (editors). *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy.* Version 2.0 (updated July 2023). Cochrane, 2023.
9. Akturk Z, Hapfelmeier A, Fomenko A, Olm M, Gehrman J, von Schrottenberg V, Teusen C, Eck S, Rehder R, Dawson S, Löwe B, Rücker G, Schneider A, Linde K. Die diagnostische Genauigkeit der 7-Item-Skala (GAD-7) und der 2-Item-Skala (GAD-2) zur Erkennung von Angststörungen bei Erwachsenen – ein systematischer Review. *Deutsche Gesellschaft für*

Die diagnostische Genauigkeit von weit verbreiteten Selbstbeurteilungs-Fragebögen zur Erkennung von Angststörungen - systematische Übersicht und Multi-Test-Metaanalyse (DIAQANDI - 01KG2105)

- Allgemeinmedizin und Familienmedizin. 57. Kongress für Allgemeinmedizin und Familienmedizin. Berlin, 28.-30.09.2023. Düsseldorf: German Medical Science GMS Publishing House; 2023. DocOS-02-02. doi: 10.3205/23degam241
10. Fomenko A, Dümmler D, Aktürk Z, Eck S, von Schrottenberg V, Teusen C, Dawson D, Rücker G, Hapfelmeier A, Linde K, Schneider A. Die diagnostische Genauigkeit des HADS-A-Fragebogens zur Erkennung von Angststörungen bei Erwachsenen – ein systematischer Review mit Metaanalyse. Deutsche Gesellschaft für Allgemeinmedizin und Familienmedizin. 57. Kongress für Allgemeinmedizin und Familienmedizin. Berlin, 28.-30.09.2023. Düsseldorf: German Medical Science GMS Publishing House; 2023. DocV-23-01. doi: 10.3205/23degam131.
  11. Fomenko A, Linde K, Aktürk Z, Schneider A, Rücker G, Hapfelmeier A. Comparison of diagnostic test accuracy estimates using the Multiple Thresholds Model based on complete and published cut-off data with a bimodal reporting pattern. Cochrane Colloquium. London, September 4 to 6, 2023. Doi: 10.1002/14651858.CD202301
  12. Linde K, Aktürk Z, Fomenko A, Eck S, Teusen C, von Schrottenberg V, Schneider A, Hapfelmeier A. Systematische Reviews mit Meta-Analyse von diagnostischen Studien verstehen und durchführen – Hands-on-Workshop. Deutsche Gesellschaft für Allgemeinmedizin und Familienmedizin. 57. Kongress für Allgemeinmedizin und Familienmedizin. Berlin, 28.-30.09.2023. Düsseldorf: German Medical Science GMS Publishing House; 2023. DocWS-36-01. doi: 10.3205/23degam320.