Abschlussbericht

für die Brandau-Laibach-Stiftung 2020

Projekt: **Kognitives Training bei gesunden Älteren – wem hilft was?**
Prognostische Modelle und Faktoren zur Vorhersage des Erfolgs von kognitiven Trainings bei gesunden älteren Erwachsenen: eine systematische Übersichtsarbeit und Metaanalyse

Antragstellerinnen: Mandy Roheger[1], M.Sc.; Anne Adams, M.Sc.[2] &

Prof. Dr. Elke Kalbe[1]

[1]Universität zu Köln, Medizinische Fakultät und Uniklinik Köln, Abteilung für Medizinische Psychologie | Neuropsychologie & Gender Studies; Centrum für neuropsychologische Diagnostik und Intervention (CeNDI)

Kerpener Str. 68

50937 Köln

Tel: +49 221 478-96244

Fax: +49 221 478-3420

[2]Institut für Medizinische Statistik und Bioinformatik

Robert-Koch-Str. 10
50931 Köln
Tel: +49 221 478-33410

E-Mail: mandy.roheger@uk-koeln.de,anne.adams@uni-koeln.de* , elke.kalbe@uk-koeln.de

Projektlaufzeit: Juli 2019 – Juni 2020

* Da die Mitantragstellerin Frau Dr. Kathrin Kuhr nicht mehr am IMSB tätig ist, hat Frau Anne Adams., M.Sc. die Projektbeteiligung übernommen.

Übersicht

Das o.g. Projekt wurde und wird von der Abteilung Medizinische Psychologie | Neuropsychologie und Gender Studies (Prof. Dr. Elke Kalbe, Mandy Roheger, M.Sc.) in Kooperation mit dem Institut für Medizinische Statistik und Bioinformatik (IMSB), Medizinische Fakultät und Uniklinik Köln (Anne Adams, M.Sc.) durchgeführt. Da die Mitantragstellerin Frau Dr. Kathrin Kuhr nicht mehr am IMSB tätig ist, wurde ihre Stelle für das Projekt von Frau Anne Adams übernommen.

Insgesamt ist zu konstatieren, dass alle avisierten Ergebnisse erreicht wurden bzw. mit insgesamt vier Publikationen mehr Ergebnisse als ursprünglich geplant erreicht werden konnten. Inhaltlich können wir mit den Arbeiten wesentlich zum aktuellen Forschungsstand „Wem hilft welches kognitive Training“ und zur Richtung zukünftiger Forschung – insbesondere hinsichtlich der Ausräumung bisheriger methodischer Limitationen in der prognostischen Forschung in diesem Bereich - beitragen.

Zur Vorbereitung des Hauptprojekts wurden drei Pilotprojekte zur Identifikation von prognostischen Modellen und Faktoren zur Vorhersage des Erfolgs von kognitiven Einzeldomänen-Trainings durchgeführt. Diese Trainings fokussieren und trainieren ausschließlich eine kognitive Domäne, z.B. das Gedächtnis. Spezifisch wurden hier systematische Übersichtsarbeiten mit Metaanalysen zu prognostischen Modellen und Faktoren zur Vorhersage des Trainingserfolgs in Gedächtnis- und Arbeitsgedächtnistrainings bei gesunden älteren Erwachsenen untersucht. Ziel war es, die Frage zu beantworten „Wem - mit welchem Profil an Ausprägungen soziodemografischer, neuropsychologischer etc. Ausprägungen - hilft Gedächtnis- bzw. Arbeitsgedächtnistraining?“.

Auf der Grundlage dieser Projekte war das Ziel des Hauptprojekts einen Schritt weiter zu gehen und den Fokus auf kognitive Multidomänen-Trainings zu legen. Diese trainieren mindestens zwei unterschiedliche Domänen (z.B. das Gedächtnis und die Aufmerksamkeit) und werden sowohl in wissenschaftlichen als auch alltäglichen Settings häufig angewendet. Im Alltag werden Multidomänen-Trainings unter anderem in kommerziell erwerblichen „Gehirn-Trainings“ eingesetzt, die in digitaler oder analoger Form verfügbar sind und insbesondere bei gesunden älteren Erwachsenen eine hohe Akzeptanz im Rahmen der Demenzprävention finden. Die Wirksamkeit von Multidomänen-Trainings konnte bereits in wissenschaftlichen Studien belegt werden, jedoch ist bisher ungeklärt, wer von dieser Art des kognitiven Trainings besonders profitiert. Ziel sollte nun die Identifikation und vergleichende Beurteilung prognostischer Faktoren und Modelle zur Vorhersage des Erfolgs von kognitiven Multidomänen-Trainings bei gesunden älteren Erwachsenen mit Hilfe eines systematischen Übersichtsartikels und einer Metaanalyse sein. Anders als in bisherigen Analysen zum Thema „Hilft kognitives Multidomänen-Training?“ ging es auch hier um die Frage „Wem hilft kognitives Multidomänen-Training“? De Arbeit soll somit zur Optimierung von

individualisierten Maßnahmen zur Stärkung der Kognition im Alter und Demenzprävention dienen. Konkret sollen die Ergebnisse auch als Grundlage für eine dann an der Uniklinik Köln durchgeführte randomisierte, kontrollierte Studie zur Überprüfung eines aus der Arbeit resultierenden Vorhersagemodells dienen.

Ergebnisse der Pilotprojekte

Ein Pilotprojekt zum Thema: „Wer profitiert von Gedächtnistraining?“ konnte bereits publiziert werden; die Publikation befindet sich im Anhang (Roheger, Folkerts, Krohm, Skoetz, & Kalbe (2020). Prognostic factors for change in memory test performance after memory training in healthy older adults: A systematic review and outline of statistical challenges. *Diagnostic and Prognostic Research*). Hier wurden soziodemographische Faktoren (z.B. Alter, Bildung, Geschlecht), (neuro-) psychologische, genetische und biologische Faktoren systematisiert, welche Veränderungen in den Domänen verbales und non-verbales Kurz- und Langzeitgedächtnis nach einem Gedächtnistraining untersuchen. Durch die Systematisierung der Faktoren und Ergebnisse der Einzelstudien konnten wir ein konsistentes, bislang nicht aufgedecktes Muster in der statistischen Berechnung von Prognosefaktoren erkennen: die vermeintlich widersprüchlichen Ergebnisse, die zu prognostischen Faktoren bislang in der Literatur beschrieben wurden, lassen sich durch unterschiedliche statistische Methoden in den Einzelstudien erklären. Denn: In den Einzelstudien werden zur Berechnung von Prognosefaktoren unterschiedliche abhängige Variablen genutzt, hierbei vor allem der Post-Test Wert (das Ergebnis eines Tests nach der durchgeführten Intervention) und der Veränderungswert (das Ergebnis eines Tests vor der durchgeführten Intervention subtrahiert von dem Ergebnis eines Tests nach der durchgeführten Intervention). Jedoch beantworten nur Rechnungen, welche den zweitgenannten Veränderungswert als abhängige Variable nutzen, die Forschungsfrage, welche uns interessiert, nämlich: Welche Individuen profitieren von einem Gedächtnistraining, haben also einen relativen Mehrgewinn? Sieht man sich die Ergebnisse hierzu an, konnten wir zeigen, dass das Alter der in Studien am häufigsten untersuchte prognostische Faktor war und dass vor allem ältere Menschen ihre Gedächtnisleistung nach einem Gedächtnistraining am stärksten verbessern konnten. Andere – weniger häufig untersuchte - Prädiktoren, zu denen signifikante Ergebnisse gefunden wurden und somit weitere Berücksichtigung in zukünftigen Studien finden sollten, waren Bildung (je weniger, desto mehr Benefit), der Persönlichkeitsfaktor „Offenheit“ (je ausgeprägter, desto mehr Benefit), Übergewicht (je weniger, desto mehr Trainingsbenefit), das genetische Merkmal ApoE 4 (Nichtträger haben mehr Benefit), sowie strukturellen Gehirnmerkmale „Integrität der weißen Hirnsubstanz“ und „Hippocampusvolumen“ (je besser bzw. größer, desto mehr Benefit) und funktionellen Hirnmerkmalen „Aktivität im frontalen Cortex bzw. Hippocampus“ (je stärker, desto mehr Benefit). Unsere Studie bietet somit wesentliche Hinweise auf die Interpretierbarkeit verschiedener Methodiken in der prognostischen Forschung sowie konkrete Hinweise auf Charakteristika, die einen Erfolg in einem Gedächtnistraining begünstigen. Vorsichtig interpretiert sieht es so aus, dass

hinsichtlich soziodemographischer Faktoren (Alter, Bildung) eher „vulnerablere“ Personen (Ältere mit weniger Bildung) profitieren. Hinsichtlich biologischer Faktoren ergibt sich das Gegenteilige Ergebnis, was bedeuten könnte, dass diese Faktoren die „Hardware“ darstellen, die eine höhere kognitive Plastizität erst ermöglichen kann. Diese Hypothesen müssen zukünftig weiter untersucht werden.

An diese Ergebnisse anknüpfend untersuchten wir in einem weiteren Teilprojekt (Roheger, Folkerts, Krohm, Skoetz, & Kalbe (under review). Prognostic models for change in memory test performance after memory training in healthy older adults: A systematic review. *Journal of Neuropsychology; Manuskript s. Anhang)* nicht nur den Einfluss von Prognostischen Faktoren, welche Veränderungen in den Domänen verbales und non-verbales Kurz- und Langzeitgedächtnis nach einem Gedächtnistraining darlegen, sondern auch den Einfluss von prognostischen Modellen. Prognostische Modelle sind definiert als meherre prognostische Faktoren, die zusammen auf einen bestimmten Outcome wirken. Anstatt also z.B. nur zu fragen: „Können jüngere Menschen mehr profitieren?“, kann man mit prognostischen Modellen mehrere Faktoren kombinieren, z.B.: „Können jüngere Menschen, die mehr Sport treiben und Gen X tragen, mehr profitieren?“. Modelle bilden eher die Lebensrealität ab, da im Menschen mehrere Eigenschaften miteinander kombiniert sind und gleichzeitig auftreten. Allerdings ist die Berechnung und Erforschung von Prognostischen Modellen auch aufgrund dieser vielen verschiedenen Möglichkeiten und Faktoren, die aufgenommen werden können, zwangsläufig komplexer. Unsere Ergebnisse zeigen, dass die Forschung zu prognostischen Modellen zu Gedächtnisverbesserungen nach einem Gedächtnistraining noch ganz in ihren Anfängen steht. Die untersuchten Studien waren methodisch teilweise unzureichend durchgeführt und dargestellt. Demnach bestand der Hauptfokus in dieser Veröffentlichung darin, die verschiedenen Arten, wie man ein prognostisches Modell rechnen kann, vorzustellen und darzulegen, damit zukünftige Forschung dies als Grundlagen nehmen können. Interessanterweise zeigten die einzig konsistenten Ergebnisse über mehrere Studien hinweg an, dass jüngere Personen, welche höher gebildet waren, am meisten in den Gedächtnistrainings profitieren konnten. Diese Ergebnisse stehen jedoch im direkten Widerspruch zu unseren Ergebnissen zu den prognostischen Faktoren. Gründe hierfür könnten sowohl in dem statistischen Unterschied zwischen prognostischen Faktoren und Modellen liegen, als auch in der Art der durchgeführten Trainings, die sich zwischen den Studien unterschieden. Es benötigt mehr Forschung, um diesem Widerspruch auf den Grund zu gehen.

In dem Teilprojekt „Wer profitiert von Arbeitsgedächtnistraining?“, zu dem ebenfalls schon ein Manuskript fertiggestellt wurde und zur Publikation angenommen wrude (Ophey, Roheger, Folkerts, Skoetz, & Kalbe (in press): Prognostic Factors of Working memory training success in healthy older adults. *Frontiers in Aging Neuroscience*), zeigt sich hinsichtlich des Faktors Alter das (scheinbar im Widerspruch zu den oben beschriebenem Muster stehenden) Ergebnis, dass vor allem die jüngeren Menschen (ab 55 Jahre) ihre Arbeitsgedächtnisleistung und Leistungen in anderen kognitiven Funktionen verbessern konnten. Außerdem sind es vor allem Menschen mit einer geringeren Leistung in den neuropsychologischen Tests zur

Baseline, die besonders von dem Arbeitsgedächtnistraining profitieren. Es scheint also zum einen so zu sein, dass es noch „Raum für Verbesserung“ geben muss in der trainierten Domäne, und zum anderen gewisse „Hardware“ Voraussetzungen vorhanden sein müssen, um von dem Arbeitsgedächtnistraining zu profitieren: So kann das Alter als eine Art Proxy für das Potential zur Neuronalen Plastizität interpretiert werden. Ist dieses Potential höher, was eher bei jüngeren Menschen der Fall ist, sind auch die Trainingseffekte bei dieser spezifischen Trainingsart größer. Zu berücksichtigen ist, dass es sich bei Arbeitsgedächtnistraining häufig um computerbasierte Testverfahren handelt, wobei die höhere Technikkompetenz jüngerer Älterer positiv wirken könnte. Schließlich wird Arbeitsgedächtnistraining zu den eher prozess-basierten Trainings gezählt, d.h. es wird eine eher abstrakte kognitive Domäne trainiert, ohne dass den Teilnehmer*innen konkrete Strategien an die Hand gegeben werden (anders als bei anderen Gedächtnistrainings). Jüngere Menschen scheinen von diesem abstrakteren Training also mehr zu profitieren, wohingegen ältere Menschen eher von einem Training profitieren, welches auf der gezielten Vermittlung von Gedächtnisstrategien basiert, wie eben den Gedächtnistrainings, die im Fokus der Studie von Roheger et al. (2020) standen.

Ergebnisse des Hauptprojekts

In unserem Hauptprojekt untersuchten wir sowohl Faktoren als auch Modelle, die Veränderungen in kognitiven Leistungen nach einem kognitiven Training vorhersagen, welches mehr als eine kognitive Domäne trainiert. Die Arbeiten sind abgeschlossen und das Manuskript in Vorbereitung (Roheger, Liebermann-Jordanidis, Krohm, Adams, & Kalbe (in preparation): Prognostic Factors and models for changes in cognitive performance after multi-domain cognitive training in healthy older adults: a systematic review).

Insgesamt screenten wir n = 10 190 Studien und konnten am Ende 23 Studien in unsere systematische Übersichtsarbeit einschließen. 13 dieser Studien untersuchten prognostische Faktoren, 10 Studien untersuchten prognostische Modelle. Es zeigte sich, dass es eine große Heterogenität zwischen den einzelnen durchgeführten Trainings gab (in der Länge, Frequenz und Dauer, aber auch in den trainierten Domänen und dem Inhalt des Trainings). Die kognitiven Funktionen, die jedoch in den meisten Trainings adressiert werden, sind Gedächtnis und Exekutivfunktionen. Untersuchte prognostische Faktoren umfassten soziodemographische Variablen (Alter, Geschlecht, Bildung), neuropsychologischer Status zu Beginn der Intervention, psychologische Variablen (z.B. Lebensqualität, depressiver Status), Trainingscharakteristiken (z.B. Intensität des Trainings), genetische Variablen, Bildgebungsparameter des Gehirns, und Marker aus Messungen mit Elektroenzephalografie (EEG). Ein relativ homogenes Ergebnis war, dass Menschen mit einem niedrigeren Baselineniveau in neuropsychologischen Aufgaben am meisten von Multi-Domänen Trainings profitieren konnten. Dies bedeutet, dass insbesondere Personen mit einer schwächeren neuropsychologischen Leistung zu Beginn des Trainings einen besonderen Nutzen aus Multidomänen-Trainings ziehen können. Zwar basiert dieses Ergebnis lediglich auf einer geringen Anzahl an wissenschaftlichen Untersuchungen, jedoch steht dies im Einklang mit

einem Ergebnis aus dem Pilotprojekt „Wer profitiert von Arbeitsgedächtnistraining?“. So liefern die Ergebnisse aus beiden Projekten Hinweise dafür, dass es einen gewissen „Raum für Verbesserung“ geben muss, damit gesunde ältere Erwachsene ihre kognitive Leistung mittels kognitiven Trainings verbessern können.

Ausblick

Unsere Ergebnisse zeigen, dass Alter sowie Bildung und kognitives Baselineniveau von neuropsychologischen Aufgaben relevante Prädiktoren für eine Verbesserung von kognitiven Funktionen nach einem kognitiven Training zu sein scheinen, wobei die Richtung der Prädiktion u.a. vom Trainingstyp abhängt. Um diese Ergebnisse zu untermauern, ist – sofern eine Finanzierung ermöglicht werden kann - geplant, eine randomisierte, kontrollierte Studie durchzuführen, in der wir diese Einflussfaktoren nochmals systematisch untersuchen.

Unsere Ergebnisse leisten außerdem einen wesentlichen Beitrag zur Aufklärung von methodischen Schwächen der bisherigen Forschungsarbeit im Bereich der prognostischen Faktoren und Modellen zur Verbesserung nach kognitiven Trainings und stellen eine Hilfestellung für zukünftige Studien dar – oder umgekehrt formuliert: leisten einen Beitrag dazu, welchen methodischen Standards Forschung zum Thema zukünftig folgen sollte. Ein weiteres Ergebnis ist, dass es wichtig wäre, eine inhaltliche Strukturierung der bislang verwendeten kognitiven Trainings vorzunehmen, um deren Wirkmechanismen genauer zu verstehen (also: welche Komponenten sind wirksam, welche weniger). Diesen wichtigen Schritt, der sich aus dem Projekt ergeben hat, setzen wir derzeit um. Eine niederländische Forschungsgruppe von der Universität Nijmegen ist auf unsere Publikationen aufmerksam geworden und kooperiert nun mit uns. Hierbei werden sämtliche Trainings klassifiziert, die wir in diesem Projekt untersucht haben, und auf ihre Wirkmechanismen hin untersucht.

Roheger et al. Diagnostic and Prognostic Research (2020) 4:7
https://doi.org/10.1186/s41512-020-0071-8

Diagnostic and Prognostic Research

REVIEW Open Access

Prognostic factors for change in memory test performance after memory training in healthy older adults: a systematic review and outline of statistical challenges

Mandy Roheger[1*], Ann-Kristin Folkerts[1], Fabian Krohm[1], Nicole Skoetz[2] and Elke Kalbe[1]

Abstract

Background: The goal is to investigate prognostic factors for change in memory test performance in healthy older adults and to report and discuss the different statistical procedures used for investigating this topic in the literature.

Methods: Prognostic factors were here understood as any measures that were investigated to estimate change in memory test performance. MEDLINE, Web of Science Core Collection, CENTRAL, and PsycInfo were searched up to November 2019. Prognostic factor and prognostic factor finding studies investigating prognostic factors on verbal and non-verbal short- and long-term memory after conducting memory training in healthy older adults were included. Risk of bias was assessed using the QUIPS tool.

Results: Our search yielded 12,974 results. We included 29 studies that address prognostic factors of change in memory test performance, including sociodemographic, (neuro-)psychological, genetic, and biological parameters. Studies showed high variation and methodological shortcomings with regard to the assessment, statistical evaluation, and reporting of the investigated prognostic factors. Included studies used different types of dependent variables (change scores vs. post-test scores) when defining change in memory test performance leading to contradictory results. Age was the only variable investigated throughout most of the studies, showing that older adults benefit more from training when using the change score as the dependent variable.

Conclusion: Overall, there is a need for adequate reporting in studies of prognostic factors for change in memory test performance. Because of inconsistencies and methodological shortcomings in the literature, conclusions regarding prognostic factors remain uncertain. As a tentative conclusion, one may say that the higher the age of the participant, the more profound the improvement in memory test performance will be after memory training.

Trial registration: CRD42019127479.

Keywords: Prognostic factors, Memory training, Prediction, Verbal memory

* Correspondence: mandy.roheger@uk-koeln.de
[1]Department of Medical Psychology | Neuropsychology and Gender Studies & Center for Neuropsychological Diagnostics and Intervention (CeNDI), Faculty of Medicine and University Hospital Cologne, University of Cologne, Kerpener Str. 68, 50937 Cologne, Germany
Full list of author information is available at the end of the article

Background

Even in the absence of severe health issues, the aging process is associated with a decline in cognitive functioning, e.g., in memory, attention, or executive functions, which may result in a loss of autonomy and quality of life in older individuals [1]. One way that has been discussed to be able to contribute to maintenance of cognitive function in the older age (> 55 years) is cognitive training (CT, defined as guided cognitive exercises designed to improve specific cognitive functions, as well as enhance performance in untrained cognitive tasks [2]). Recent meta-analyses and reviews show that CT can be effective not only in improving cognitive functions in healthy older individuals, but also their quality of life [3, 4]. There are many different types of CT, which differ regarding their settings (e.g., single vs. group settings), materials used (e.g., computerized vs. paper-and-pencil tasks), but also regarding their focus on different outcomes (e.g., memory, attention, executive functions). Memory, which is a key function that typically decreases in higher age, even in healthy older adults [5], can also be improved or maintained with the help of CT [4]. However, one question that remains underinvestigated is: who (with which profile of, e.g., sociodemographic, neuropsychological, genetic parameters) benefits from CT? Yet, identifying prognostic factors is highly important for providing new treatment options and in term of dementia prevention [6]. Prognostic factors (in literature also often referred to as "predictors") for changes in test performance after a CT that are under debate are sociodemographic variables, brain imaging parameters, genetic parameters, and blood factors, as well as personality traits, cognitive and non-cognitive abilities at the entry of the training, and different training characteristics, e.g., intensity of the trainings [7]. Yet, data is highly inconsistent: for example, there are several studies that report higher age as a positive prognostic factor for changes in test performance after a CT in healthy older adults [7, 8], while some studies indicate that younger individuals show improvement in test performance after a CT [9, 10].

Yet, inconsistent results regarding prognostic factors of CT can be seen throughout the prognostic factor literature for CT benefits so far, and the question arises, why this is the case. Until now, no systematic review exists investigating prognostic factors for CT success in healthy older adults in general, and memory training in particular to answer this question [11]. However, considering the fact that prognostic factors for change in cognitive performances after a CT in healthy older adults have many potential uses (e.g., aiding treatment and lifestyle decisions, improving individual dementia risk prediction, providing new treatment options [6]), and data so far reveals highly inconsistent results, systematic reviews and meta-analyses are urgently needed to summarize evidence about the prognostic value of particular factors to help to match cognitive interventions to individuals to improve their effectiveness in regard of a personalized medicine approach [12, 13].

Therefore, the present review focuses on prognostic factors for changes in memory performances after memory training, due to different reasons: first, memory belongs to the most vulnerable cognitive functions in aging (e.g., [5]). Second, we wanted to get a first overview over the published data on prognostic research after training interventions in a narrower frame, therefore focusing only on one specific relevant domain. Conclusions from this review could then help further research on prognostic factors of cognitive change induced by CTs.

Objectives

The main goal of the present systematic review is to investigate prognostic factors for changes in memory performance after memory training in healthy older adults. Further, we wanted to investigate different methods used to evaluate prognostic factors for changes in memory performance after memory training. Based on the checklist for critical appraisal and data extraction for systematic reviews of prediction modelling studies [12, 14, 15], which can also be used to assess prognostic factors studies [12], we defined our systematic review question using the "PICOTS system" [15]. Our target population are healthy older individuals, defined as individuals aged ≥ 55 years with absence of any neurological or psychiatric disease (P). Regarding the investigated intervention (I), we investigated all prognostic factors assessed for change in memory test performance after memory training. No comparator factor is being considered (C). Outcome events for this review are changes in memory test performance after memory training in the domains verbal short-term memory, verbal long-term memory, as well as non-verbal short- and long-term memory operationalized with objective and standardized measurement instruments (O). The measurement of the prognostic factor had to be done before the memory training started and all follow-up information on the outcomes (all time periods) was extracted from the studies (T). Finally, prognostic factor measurement was studied in non-clinical settings to provide prognostic information for possibilities of prevention of cognitive decline (in other words, possibilities to strengthen cognitive function) in cognitively intact individuals (S).

Methods

The present systematic review was preregistered; the review protocol can be assessed at www.crd.york.ac.uk/

PROSPERO/ (ID: CRD42019127479). The reporting follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guideline for systematic reviews and meta-analysis [16]. "The PRISMA for Abstracts Checklists", as well as "The PRISMA checklist for systematic reviews" are displayed in Supplementary Tables 1 and 2.

Search and study selection

A systematic search was conducted in MEDLINE Ovid, Web of Science Core Collection, CENTRAL, and PsycInfo up to October 2018. An update-search was conducted in the same data bases until 12th November 2019. Reference lists of all identified trials, relevant review articles, and current treatment guidelines were hand searched for further literature. In cases where no full text could be obtained, we contacted the authors and asked them to provide full text publications within a 2-week time frame. Further information on the systematic search and the full search strings for each database are presented in the Supplementary Material, Tables 3, 4, 5 and 6.

Titles and abstracts were screened according to predefined eligibility criteria by two individual review authors (MR and AKF) with the Covidence Software (Veritas Health Innovation) [17]. Afterwards, the full-text articles of the studies meeting the inclusion criteria were further reviewed for inclusion in the systematic review. In cases where no consensus could be reached between the two authors MR and AKF, a third author (NS) was asked and the case was discussed until a final consensus was reached.

Eligibility criteria

The review focused on peer-reviewed studies in English and German with no limitations regarding publication date which investigated prognostic factors of changes in memory test performance after memory training. Full study reports needed to be available; abstracts, books, book chapters, study protocols, and conference papers were excluded.

Prognostic factor studies on healthy older participants (age ≥ 55 years) were included. Data from participants with dementia diagnosis, neurological and/or psychiatric diseases, as well as uncorrected seeing or hearing impairments, assessed at least via self-report, were excluded. Studies with participants with mild cognitive impairment (if reported) were also excluded as we want to investigate healthy adults in the context of interventions.

Regarding the investigated intervention and included prognostic factors, all prognostic factors (e.g., sociodemographic factors, brain imaging parameters, genetic parameters, blood factors, personality traits, cognitive abilities at the entry of the training, different training characteristics, e.g., intensity of the trainings, etc.) which investigate changes in memory test performance after memory training were included in the review and meta-analysis. Memory training was defined as a CT that targets primarily on memory performance with a minimum of two sessions in total. The memory training can either include computerized or paper-pencil tasks with clear cognitive rationale, which are administered either on personal devices or in individual- or group settings held by a facilitator. When multi-domain approaches were examined, memory had to be the main component of the program (at least 50% of the exercises).

Prognostic factor studies, which investigate memory training benefits as an outcome (verbal or non-verbal short- or long-term memory) measured with established objective neuropsychological tests, were included. Working memory was excluded and is being investigated in a different review, as we define working memory as an executive function rather than a pure memory function [18]. We excluded subjective self-rated memory scales, as well as measures of memory strategy use. The factor measurement of the included studies had to be conducted before the memory training started, and there was no limitation regarding the length of the follow-ups.

Data extraction

Two review authors (MR and AKF) independently extracted the data according to the Critical appraisal and data extraction for systematic reviews of prediction modelling studies_ prognostic factors (CHARMS_PF) checklist [15] to investigate the reporting of prognostic factors.

Quality assessment

Two reviewers (MR and AKF) independently assessed the extracted studies for the risk of bias using the Quality in Prognosis Studies (QUIPS) checklist, developed by Hayden et al. [19] to examine the risk of bias in prognostic factors studies across six domains [19]: Study participation, study attrition, prognostic factor measurement, outcome measurement, adjustment for other prognostic factors, statistical analyses, and reporting. Each of the six domains was judged with high, moderate or low risk. A detailed description of the domains included in the tool and the judgment taken by the two reviewers is presented in Supplementary Material 7.

Statistical analyses

In the pre-registration of the study, we registered a meta-analysis to investigate the predictive performance of the different prognostic factors. The goal was to meta-analyze groups of "similar" prognostic effect measures with a random effects approach to allow for unexplained heterogeneity across studies. However, after the data extraction, we found that data on prognostic factors

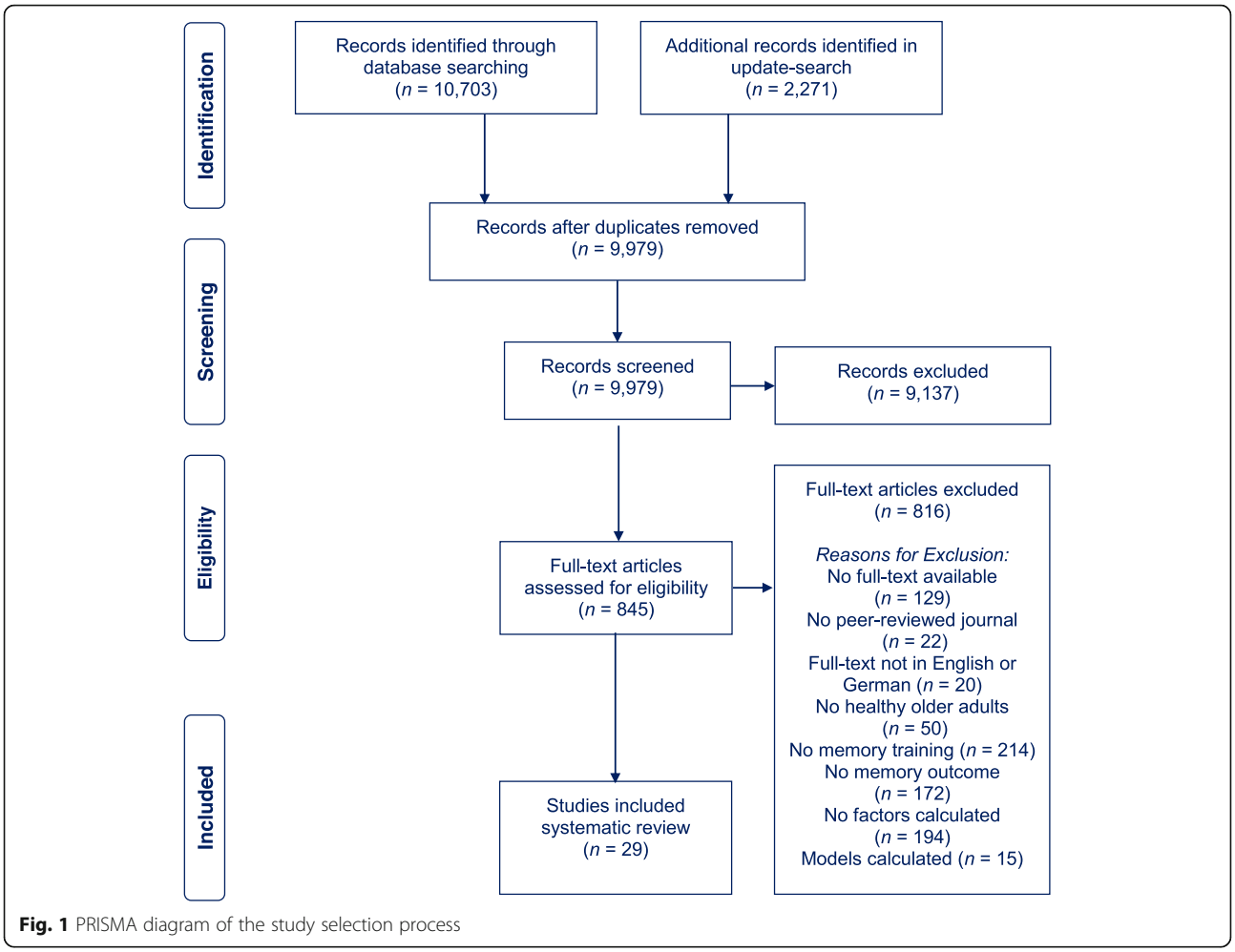


Fig. 1 PRISMA diagram of the study selection process

of changes in memory test performance after memory training were too heterogeneous and too poorly reported to conduct a meta-analysis.

Results

Study selection

The total number of retrieved references and the numbers of included and excluded studies with reasons for exclusions are documented in a flow chart as recommended in the PRISMA statement [16]. The PRISMA diagram in Fig. 1 illustrates the study selection process. Further, 10,703 studies were identified through the database search and by scanning the included studies in previously published systematic reviews and meta-analysis on memory training success in healthy older adults, n = 2271 studies were identified in an update search. After removing the duplicates, n = 9979 studies were screened. It was difficult to distinguish, from study abstracts alone, between prognostic factor finding studies and model development studies. We thus assessed 845 full-texts for eligibility. Finally, n = 29 studies were included in the present review. All studies were published in English.

Data extraction

A main challenge was to distinguish between prognostic factor finding and model development studies, as the authors in general did not state their aim regarding prognostic factors or models. Therefore, we used full text interpretations to classify studies as prognostic factor finding or model development studies. Eight discrepancies were resolved after discussion with a third reviewer (NK) with experience in the field of prognostic research.

Study characteristics

An overview of the main characteristics of the included studies is outlined in Table 1. Further information of the included studies is illustrated in Supplementary Tables 8 and 9.

Of the 29 studies included, we found that 15 studies used a randomized controlled design, whereas six studies only used a controlled design (Table 1). Furthermore, eight studies used a non-randomized, non-controlled longitudinal study design, which may be classified as a cohort study, as the defining characteristic of the cohort

Table 1 Study and participant characteristics of the included studies

Study	Sample				Training		Outcomes	Prognostic factors
Study design Initial sample size for the experimental group Dropout and reasons	Age (years, M, SD)	Sex ♂	 ♀	Education(years, M, SD)	Description of memory training—content and frequency	Total length of training in minutes	Definition and method of assessment Timing of outcome assessment	Definition and methods
Pesce et al. [20] Stratified randomized study *n* = 30 *n* = 29	70.40 (7.00)	14	15	9.60 (1.80)	Method of loci and general strategies. 24 weeks, 2 times a week for 1 h	2880	RAVLT, MMSE	Antioxidant levels assessed with the Biological Antioxidant potential Test; reactive oxygen metabolites derivative compounds assessed with the d-ROMs Test
O'Hara et al. [9] Non-randomized, non-controlled longitudinal study *n* = 531 *n* = 419 due to several reasons at 5-year follow-up	73.73(7.62)	34	78	15.56 (2.79)	Method of loci. 2 weeks, 5 times a week for 2 h	1200	Number of words correctly recalled, number of words correctly recalled in order. Assessed at baseline and 5-year follow-up measurement	Pre-training, gain scores following training, age, education, reported use of mnemonic at follow-up, type of pre-training (standard vs. comprehensive) and length of training.
Mohs et al. [21] RCT *n* = 68 *n* = n.a.	78.30 (7.40)	15	53	16.00 (2.70)	Structured memory training focusing on memory improvement and different strategies. Nine 90-min sessions	810	Verbal memory assessed with CVLT, non-verbal memory assessed with BFLT. Assessed 2 times at baseline, at post-test, 3 months and 6 months follow-up	Age, education, gender, subjective reported memory assessed with the MFI and the MFQ
Kirchhoff et al. [22] Non-randomized, non-controlled longitudinal study *n* = 16 *n* = 2 due to technical difficulties	72.00 (66–81)	7	7	14.70 (2.90)	Memory strategy training and practice. 2 training sessions	Missing information	Memory retrieval using Remember/Know/New recognition memory decisions Assessed at pre-training and post-training	Hippocampal activity
Kirchhoff et al. [23] Controlled trial *n* = 16 *n* = n.a.	71.9 (66–81)	8	8	14.8 (2.7)	Memory strategy training and practice 2 training sessions.	Missing information	Recognition memory using Remember/Know/New recognition memory decisions. Assessed at pre-training and post-training	Activity in prefrontal cortex, left lateral temporal cortex.
Leahy et al. [24, 25] RCT *n* = 22 *n* = 1	74.77 (6.57)	8	13	18.77 (2.62)	Memory specificity training to improve the specificity of older adults' retrieval of autobiographical memories by providing systematic practice. 4 weeks, once a week for 60 min	240	Autobiographical memory specificity. Assessed at pre-test, post-test, and 3 months follow-up.	Memory specificity assessed with MEPS, functional limitations assessed with FLP, self-rated depression assessed with HADS, independence assessed with IADL
Andrewes et al. [26] RCT controlled for sex *n* = 20 *n* = 3	60–70 years	10	10	Some secondary schooling: *n* = 3 Secondary school + trade qualifications: *n* = 5 Complete secondary school: *n* = 6	Memory handbook training for face-name and prospective memory areas; independently implemented at home 4 weeks, 30 min per session	Missing information	Improvement in: Face-name Test, Laboratory Prospective Memory Assessment, Everyday Prospective Memory Assessment. Assessed at pre-test, post-test and 4-month FU	RAVLT, Warrington Forced-Choice Recognition for Faces, BDI, NART, Mattis Dementia Rating Scale.

Table 1 Study and participant characteristics of the included studies *(Continued)*

Study	Sample				Training		Outcomes	Prognostic factors
				Began tertiary school: *n* = 6				
Anschutz et al. [27] Non-randomized, non-controlled longitudinal study *n* = 10 *n* = 1 due to severe illness	73.50 (n.a.)	2	7	10.70 (n.a.)	Method of loci No information on training duration and frequency	Missing information	Free-recall of two lists and recognition of two lists consisting of 12 nouns each. Assessed at pre-test and 34 months after finishing the training	Free-recall pre-test, free recall list 1, age
Bissig and Lustig [28] Non-randomized, non-controlled longitudinal study *n* = 19 *n* = 1 due to low accuracy of studied words	74.50 (6.10)	n.a.	n.a.	18.00 (3.30)	Modified recollection training procedure 2 weeks, 4 sessions per day at 7 days	Missing information	Ranking: participants were ranked by final lag level (lag between lure repetitions). Assessed and adapted individually during each training performance	Age, crystallized intelligence
Bråthen et al. [29] Controlled trial *n* = 126 *n* = 3	Old: 73.40 (3.00)	Old: 29	Old: 49	Old: 14.70 (2.90)	Learning and practicing the Method of loci technique aiming to improve episodic memory performance 10 weeks, once a week + 8 weekly online home assignments	Missing information	Memory improvement: change in correct written recall of word list consisting of 100 nouns. Assessed at pre-test and post-test	Cortical volume, hippocampal volume, ALFF, fALFF
Brooks et al. [8] RCT *n* = 224 Dropout not reported	68.58 (7.05)	n.a.	n.a.	15.33 (2.58)	Pre-training: imagery training, verbal elaboration and relaxation. Name-Face Mnemonic: three-step mnemonic Method of loci: method of loci for serial word recall. 2 weeks, 5 times a week for 120 min)	1200	Proper name recall task, word recall task (16 common words). Assessed at pre-test and post-test	Pretraining, pretest score, age, length of training, pretraining x length
Clark et al. [30] Multi-site RCT (ACTIVE) *n* = n.a. *n* = n.a.	No demographics separately for the memory training groups were reported.				Memory training focused on improving verbal episodic memory through instruction and practice in strategy use 6 weeks, 10 60-min sessions	600	HVLT, RAVL, RBMT. Measured at baseline, immediate post-training, 1-;2-, 3-; 5-, and 10-year FU	Obesity, determined from BMI (in kg/m^2) computed from measured height and weight data obtained at baseline
Clark et al. [31] Multi-site RCT (ACTIVE) *n* = n.a. *n* = n.a.	No demographics separately for the memory training groups were reported.				Memory training focused on improving verbal episodic memory through instruction and practice in strategy use 6 weeks, 10 60-min sessions	600	Hopkins Verbal Learning Test, RAVL, RBMT. Measured at baseline, immediate post-training, 1-;2-, 3-; 5-, and 10-year FU	Education (self-reported as years of completed schooling)
de Lange et al., [32] Controlled trial *n* = 76 *n* = 9 due to time constraints	73.60 (3.00)	25	51	15.00 (2.70)	Learning and practicing the Method of Loci technique aiming to improve episodic memory performance. 10 weeks, once a week + 8 weekly online home assignments	Missing information	Memory improvement: Word list recall	Interindividual variability in white matter microstructure

Table 1 Study and participant characteristics of the included studies *(Continued)*

Study	Sample				Training		Outcomes	Prognostic factors
de Lange et al., [33] Controlled trial *n* = 44 *n* = 0	73.30 (2.70)	21	23	15.70 (3.10)	Learning and practicing the Method of Loci technique aiming to improve episodic memory performance. 10 weeks, once a week + 8 weekly online home assignments	Missing information	Memory improvement: word list test (100 words)	White matter microstructure
Tomaszewski Farias et al. [34] Multi-site RCT (ACTIVE) *n* = n.a. *n* = n.a.	No demographics separately for the memory training groups were reported				Memory training focused on improving verbal episodic memory through instruction and practice in strategy use 6 weeks, 10 60-min sessions	600	Memory factor: Immediate recall HVLT, RAVLT, paragraph recall, RBMT	Instrumental activities of daily living, 18 questions of the Minimum Dataset Home Care scale
Finkel and Yesavage [35] Controlled trial *n* = 77 *n* = 16 due to illness (*n* = 5), frustration (*n* = 7), bad weather (*n* = 2), no reason (*n* = 1)	71.29 (6.31)	30%	70%	n.a.	Method of loci No information on training duration and frequency	Missing information	Memory improvement gain scores of a list of 16 common words recall	Age, education, MMSE score, depression score, neuroticism and extraversion scale of the NEO-PI
Hampstead et al. [36] RCT *n* = 12 *n* = 1 due to ongoing disease	73.20 (7.70)	n.a.	n.a.	16.10 (3.40)	Object Location Assignment encoding and retrieval with mnemonic strategy from a cognitive rehabilitation program 2 weeks, 5 sessions + 1 follow-up session one month later	Missing information	Modified change score of Object Location Assignment accuracy	Medial temporal lobe volumetrics (hippocampus, amygdala, inferior lateral ventricles), standardized neuropsychological measures (RBANS Delayed Memory Index, TMT B)
Hill et al. [37] Controlled trial *n* = 59 *n* = n.a.	67.80 (7.50)	n.a.	n.a.	5.80 (1.10)	Mnemonic training 2 weeks, twice a week for 120 min	1680	Recall performance in name-face recall	Rated confidence (perceived confidence in recalling the names of unfamiliar faces).
Hill et al. [38] Non-randomized, non-controlled longitudinal study *n* = 102 *n* = n.a.	75.40 (10.50)	32	70	n.a.	Name- and face and list-learning program using an imagery and judgment technique and method of loci method. 2 weeks, 7 times a week for 120 min	1680	Improvement in Name-Face recall, Improvement in List-Recall	MMSE.
Leahy, Ridout, and Holland, [24] RCT *n* = 20 *n* = 1 due to unrelated health problems	76.85 (5.27)	6	14	17.75 (2.65)	Memory flexibility program 4 weeks, once a week for 60 min	240	Autobiographical memory specificity in the AMT. Assessed at pre-test, post-test, and 3 month FU.	Baseline cognitive flexibility measured with the verbal fluency sub-score of ACE-III.
López-Higes et al. [39]	ApoE 4	n.a.	n.a.	n.a.	Memory training consisting	2700	Logical Memory and Word	Apolipoprotein E genotyping

Table 1 Study and participant characteristics of the included studies *(Continued)*

Study	Sample				Training		Outcomes	Prognostic factors
RCT *n* = 50 *n* = 0	carriers: 71.64 (5.72) Non-carriers: 71.68 (5.65)				of cognitive stimulation, memory concepts, management of forgetting everyday experiences, meta-memory training 3 months, 30 90-min sessions		List from WMS-III	
McDougall et al., [40] RCT *n* = 135 Loss to post-test: *n* = 8 Loss to FU: *n* = 12 Loss to end of study: *n* = 8	74.69 (5.74)	30	105	13.39 (3.90)	CBMEM-based intervention, based on the four components of self-efficacy theory 4 weeks, twice a week including 8 sessions and 4 booster sessions	720	HVLT-R, BVMT-R, RBMT. All outcome measures were administered at baseline, post-class (2 months after baseline), post-booster (6 months), post-classroom FU (14) and at the end of study (24 months)	Ethnicity, group assignment, time, education
McDougall et al., [41] RCT *n* = 135 Loss to post-test: *n* = 8 Loss to FU: *n* = 12 Loss to end of study: *n* = 8	74.69 (5.74)	30	105	13.39 (3.90)	CBMEM-based intervention, based on the four components of self-efficacy theory 4 weeks, twice a week including 8 sessions and 4 booster sessions	720	Relative gains in HVLT-R, RBMT All outcome measures were administered at baseline, post-class (2 months after baseline), post-booster (6 months), post-classroom FU (14) and at the end of study (24 months)	Age, education, racial/ethnic group
Neely & Bäckman [42] RCT *n* = 23 *n* = n.a.	73.00 (4.20)	4	19	9.90 (3.10)	Encoding operations including interactive imagery and method of loci; attention training, relaxation training. Training was conducted in groups with 11–12 subjects, met twice a week for 5 consecutive weeks, each session lasted 1.5 h	900	Recall of concrete words, recall of objects, recall of subject-performed tasks, recall of abstract words Assessed at pre-test, post-test directly after training, 6 months FU	Pretest score for each dependent variable, MMSE score, age, years of education
O'Hara et al., [43] Non-randomized, non-controlled longitudinal study *n* = 212 *n* = 113	74.00 (7.90)	68	32	15.50 (2.70)	Memory training was not further described. Missing information on duration and frequency.	Missing information	BVRT, Logical Memory Test, Associate Learning Test, List-learning test. Assessed at baseline and FU 4-5 years after memory training.	Apolipoprotein E genotyping.
Park et al. [7] RCT *n* = 39 *n* = n.a.	69.81 (4.90)	11	28	11.41 (4.31)	Multi-strategic memory training. 10 sessions once a week, each session lasted 1.5 h	900	Elderly verbal learning test of the EMS to assess verbal memory; Simple Rey Figure Test of the EMS to assess non-verbal memory. Assessed at pre-test and post-test (within 3 months after finishing the training)	All baseline values of the scores of neuropsychological tests, age, gender, years of education
Rosi et al., [44] Non-randomized, non-	68.73 (6.05)	n.a.	n.a.	11.36 (3.50)	Memory training program. 6 weeks, once a week for	360	Word list learning (memory practiced task), grocery list	Vocabulary test, Raven standard progressive matrices, listening

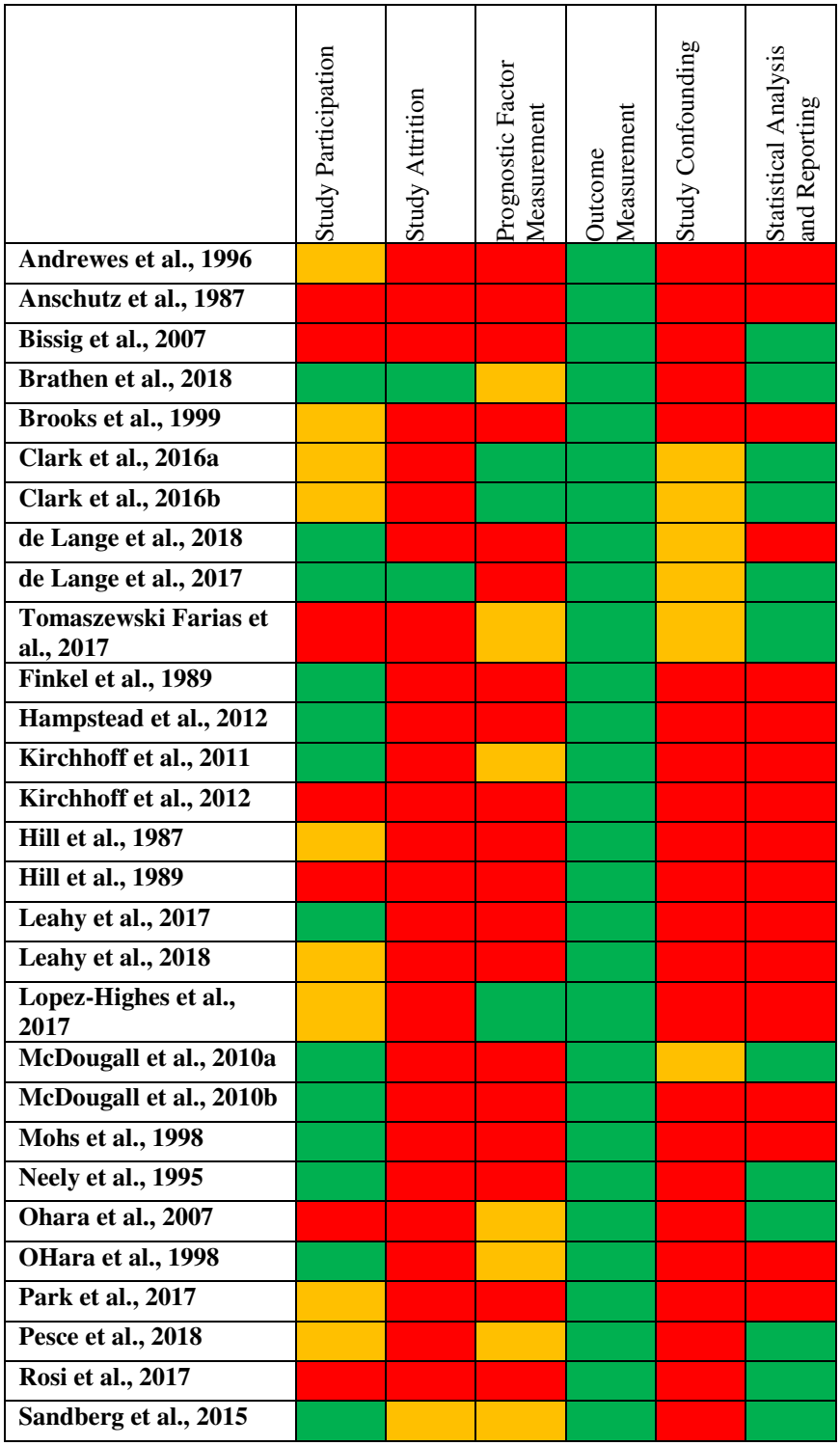
Table 1 Study and participant characteristics of the included studies *(Continued)*

Study	Sample				Training		Outcomes	Prognostic factors
controlled longitudinal study *n* = 44 *n* = n.a.					60 minutes.		learning (memory non-practiced task), associative learning Assessed at pre-test and post-test.	span test, letter comparison, age
Sandberg et al. [45] Non-randomized, non-controlled longitudinal study *n* = 112 *n* = 18 due to various reasons	70.90 (6.70)	38	56	11.90 (3.70)	Mnemonic training was based on the Swedish version of the number-consonant mnemonic task 5 times, twice a week	600	Number recall. Assessed at pre-test, post-test and FU	Three measures of episodic memory (free recall of concrete nouns, free recall of abstract nouns, paired-associate recall), three measures of working memory (listening span, two versions of computation span), nine measures of processing speed, two measures of verbal knowledge, depression (ZSRDS), vocabulary

All reported values regarding sample size, dropouts, and sociodemographic variables only refer to the memory training groups. For the variables age (in years) and education (in years) means and standard deviations were displayed, when reported. Otherwise, ranges and/or absolute numbers are stated

RAVLT Rey Auditory Verbal Learning Task, *BDI* Beck Depression Inventory, *NART* National Adult Reading Test, *dROMs* reactive oxygen metabolites derivative compounds, *FU* follow-up, *ALFF* amplitude of low-frequency fluctuation, *fALFF* fractional amplitude of low-frequency fluctuation, *BMI* body mass index, *MMSE* Mini-Mental State Examination, *NEO-PI* NEO Personality Inventory, *RBANS* Repeatable Battery for the Assessment of Neuropsychological Status, *TMT B* Trial Making Test Version B, *AMT* Autobiographical Memory Task, *ACE-III* Addenbrooke's Cognitive Examination-III, *CBMEM* Cognitive Behavioral Model of Everyday Memory, *HVLT-R* Hopkins Verbal Learning Test-Revised, *RBMT* Rivermead Behavioral Memory Test, *BVMT-R* Brief Visuospatial Memory Test revised, *EMS* Elderly Memory Disorder Scale, *BVRT* Revised Benton Visual Retention Test, *WMS-III* Wechsler Memory Scale III, *HVLT* Hopkins Verbal learning task, *MEPS* means end problem solving procedure, *FLP* functional limitation profile, *HADS* Hospital Anxiety and Depression Scale, *IADL* instrumental and basic activities of daily living, *CVLT* California Verbal Learning Test, *BFLT* Biber Figure Learning Test, *MFI* memory controllability inventory, *MFQ* Memory Functioning Questionnaire, *ZSRDS* Zung Self-Rating Depression Scale

Table 2 Risk of bias assessment



Red color indicates a high risk of bias, yellow color indicates a medium risk of bias, green color indicates a low risk of bias, assessed with the QUIPS tool [18]

is the participants' health status and attendance in memory training.

The sample sizes of the memory training intervention groups varied greatly between the studies, ranging from n = 10 participants [27] to n = 531 participants [9], with three studies not giving clear information on how many participants attended the memory training [30, 31, 34].

A detailed description of the different memory training interventions used (regarding content, length, and frequency) is displayed in Table 1. Seven studies stated that a strategy CT using the Method of Loci was conducted [8, 9, 27, 29, 32, 33, 35]. All other training programs differed in their content (e.g., learning and practicing of different memory strategies, memorizing grocery lists, psychoeducation about memory processes).

The mean age of the samples ranged from 67.8 years [37] to 78.3 years [21]. Yet, the samples were highly educated throughout the studies, ranging from a mean of 11.9 years [45] to a mean of 18.77 years of education [24, 25]. The mean score on the Mini-Mental State Examination (MMSE), which was assessed in 13 studies at baseline as an indicator for the participant's global cognitive status at baseline, ranged from a mean of 25.9 points [30, 31] to 29.2 points [44]. In most studies, the samples consisted of more women than men, with an overall of 65.9% women and 34.1% men participating in the studies.

Risk of bias

Regarding the reporting quality, Table 2 shows the risk of bias assessment according to the QUIPS tool [19] in all included studies. The table shows that there is important information lacking, especially regarding the domains study attrition, prognostic factor measurement, study confounding, and statistical analysis and reporting. Interestingly, the parameter outcome measurement was the only one in which all 29 studies provided a sufficient reporting and were rated as having a low risk of bias. A further important result was that statistical analysis and reporting was correctly accounted in eleven studies [9, 28–31, 33, 34, 40, 42, 44, 45]. Yet, all other studies which used correlation analysis or group comparisons as statistical methods to quantify prognostic factors were rated with a low reporting quality. This was also the case if no data was provided. Overall, the reporting quality was in part insufficient, and the studies in their entirety were difficult to comprehend, especially regarding the prognostic factor measurement, confounding and statistical analysis.

Outcomes and statistical outcome measures

In the present review, we investigated four outcomes: verbal short-term memory, verbal long-term memory, non-verbal short-term memory, and non-verbal long-term memory. Outcomes were well defined in all investigated studies. However, only five studies [7, 24, 25, 36, 42] reported that they blinded the outcome measurement. For a detailed overview of the different outcomes and their assessment, see Tables 3, 4, 5, and 6.

Twenty-one out of the 29 studies investigated verbal short-term memory as an outcome. Seven studies [29, 32, 33, 35, 39, 42, 44] used the immediate recall of a word list, which was the most frequently used test in this domain.

Twelve out of the 29 studies investigated verbal long-term memory. The delayed recall of a word-list test was the most frequently used test in four studies [9, 27, 38, 43].

Non-verbal short-term memory was only assessed in two out of 29 studies: one study used the immediate recall of the Simple Rey Figure test [7], the other used the Biber Figure Learning Test [21].

Four out of 29 studies assessed non-verbal long-term memory, all of them using different tests as outcome measures (see Table 6).

Prediction of more than one outcome was common, which may be due to their mostly exploratory aim.

Not only the used tests to measure the outcomes differed, but there was also substantial heterogeneity in the statistical outcome measures used. In total, eight studies used the post-test scores as the dependent variable for their calculations, whereas 18 studies used the change score (defined as post-pre scores) as the dependent variable for their prognostic factor calculation. Residual change scores were used as the dependent variable in only four studies, all of the defined as an outcome in the domain verbal short-term memory [32–34, 37]. For nine outcomes, there was no clear definition of the dependent outcome variable used for the prognostic factor measurement. None of the studies used percentile change scores as the dependent variable.

Prognostic factors and statistical methods of prognostic factor analysis

There was no detailed description (e.g., a separate paragraph stating not only the name of the prognostic factor and method of measurement, but also blinding, and use in the statistical analysis (e.g., as a continuous or dichotomous factor)) of the candidate prognostic factors in most of the studies. Investigated prognostic factors include sociodemographic variables (i.e., age, sex, education, and ethnicity), neuropsychological test status at study entry in different domains, brain imaging measures, genetic variables (i.e., apolipoprotein E4), training characteristics, and personality traits (for a detailed overview, see Tables 3, 4, 5, and 6). The prognostic factor neuropsychological status at study entry, examined in 13 studies, was the most assessed prognostic factor [7, 8, 24–26, 28, 35, 38, 41–45], followed by age, which was assessed in eleven studies [7, 8, 21, 28, 35, 40–45]. Concerning other sociodemographic factors, education was tested as a prognostic factor in nine studies [7, 9, 21, 30, 31, 35, 40–42]; sex, however, was only investigated in two studies [7, 21] as a prognostic factor for changes in memory test performance after memory training. Six studies investigated different imaging factors [22, 23, 29, 32, 33, 36]. Other investigated prognostic factors were ethnicity [40, 41], subjective reported memory [21], depression [26, 35], "BIG 5"

Table 3 Prognostic factors for training improvement in verbal short-term memory

Study	Test for outcome assessment	Dependent variable	Prognostic factor					
Multiple regression								
			Age	Education	Sex	Neuropsychology	Imaging	Others
de Lange et al., [32]	Word list	Standardized residuals					White matter microstructure →	
McDougall et al. [40]	HVLT RBMT	Relative gains	↑			Pre-test score ↑		Ethnicity →
Neely and Bäckman [42]	Immediate recall of word list	Post-test scores	↓	↑		MMSE ↑ Pre-test score ↑ *		
Rosi et al. [44]	Immediate recall of word list	Post-test scores	↓			Pre-test ↑* Working memory ↓ Fluid ability ↓ Crystallized ability ↑* Processing speed ↑ Short-term memory ↓		
Sandberg et al. [45]	Number recall	Post-test scores	↓*			Episodic memory ↑* Processing speed ↓ Working memory ↑* Verbal knowledge ↑		
Brooks et al. [8]	Name recall	Post-test scores	↑*			Pre-test score*		Pretraining x mnemonic training →
Correlation analysis								
Mohs et al. [21]	HVLT	Post-test scores	→	→	→			Subjective reported memory →
Kirchhoff, Anderson, Smith, Barch et al., [22]	Recognition memory decisions	Change score					Activity in frontal cortex ↑	
Kirchhoff, Anderson, Smith et al., [22]	Recognition memory decisions	Change score					Activity in hippocampus ↑	
Andrewes et al. [26]	Face-name test	Change score				NART → RAVT → Warrington Forced Choice Recognition ↑		Depression → Mattis dementia s cale →
Bråthen et al. [29]	Immediate recall of word list	n.a.					Hippocampal volume ↑* Amplitude of low frequency fluctuation ↓ Fractional amplitude of low frequency fluctuation ↓*	
Finkel and Yesavage [35]	Immediate recall of word list	Gain scores	x	x		MMSE x		Openness of experience ↑* Depression x Extraversion x Neuroticism x
Hill et al. [37]	Face-name recall	Standardized residual scores						Rated confidence ↑

Table 3 Prognostic factors for training improvement in verbal short-term memory *(Continued)*

Study	Test for outcome assessment	Dependent variable	Prognostic factor			
Hill et al. [38]	Face-name recall	Performance changes		MMSE ↑		
Group comparisons (ANOVA, *t* test)						
Clark, Xu, Callahan et al., [30]	HVLT RAVL RBMT	Relative mean improvement				Obesity ↓*
Clark, Xu, Unverzagtet al., [31]	HVLT RAVL RBMT	Relative mean improvement	→			
McDougall et al. [40]	HVLT RBMT	n.a.	↓			Ethnicity (Blacks and Hispanics scored lower than Whites)
Mixed models						
Tomaszewski Farias et al. [34]	HVLT RAVL RBMT	Normalized residuals				Activities of daily living ↑
López-Higes et al. [39]	Word list recall Logical memory test	n.a.				Apolipoprotein E4 →
No clear reporting						
Bissig and Lustig [28]	Rank-test	n.a.	↓	Crystallized intelligence ↑		
de Lange et al., [33]	Word list	Standardized residuals			White matter microstructure ↑	

Studies are sorted according to the statistical method used for obtaining the prognostic factors
HVLT Hopkins Verbal learning Task, *MMSE* Mini Mental State Examination, *NART* National Adult Reading Test, *RAVL* Rey Auditory Verbal Learning Test, *RBMT* Rivermead Behavioral Memory Test, ↑ the higher the prognostic factor, the higher the improvement/positive correlation, ↓ the lower the prognostic factor, the higher the improvement/negative correlation, → no direction of effect reported, * significant, x unclear reporting

personality traits [35], self-rated confidence [37], obesity [30, 31], activities of daily living [24, 25, 34], apolipoprotein E 4 (a protein that is involved in the fat metabolism of the body and constitutes a risk factor for Alzheimer's disease) [39, 43], biological antioxidant potential [20], and length of memory training [8, 9].

There were several different statistical methods used to calculate the impact of prognostic factors after memory training on memory outcomes. Eight studies calculated a multiple regression [7–9, 32, 41, 42, 44, 45] and two studies used a mixed model approach [34, 39]. Notably, 12 studies used correlation analysis to investigate prognostic factors [21–27, 29, 35–38]. Four studies [30, 31, 40, 43] used group comparisons (e.g., ANOVAs, *t* tests). In two studies [28, 33], there was no clear reporting on which statistical methods were used to determine the prognostic factors.

Prognostic factors of change in memory test performance after memory training

One of the overall aims of the present systematic review was to systematize which prognostic factors are predictive for which of the four investigated memory outcomes. The results are summarized in Tables 3, 4, 5, and 6, structured according to the statistical method used for calculating the prognostic factors and the dependent outcome variables. There is a similar pattern that can be detected over all four outcome domains: The direction of the relationship between the prognostic factor and the memory outcome (the more of x/ the less of x) differ depending on which dependent variable is evaluated as the outcome measure. This finding is substantial for the interpretation of the current literature on prognostic factors of changes in memory test performance after memory training in healthy older adults.

The prognostic factor *age* was the factor that was investigated in most studies. Studies that used the post-test scores as the dependent outcome measure showed that participants with lower age showed greater improvements in memory test performance after training [9, 42, 44, 45] with only one exception [8]. However, it should be noted that the study of Brooks et al. [8] also integrated an interaction term in their analysis. In contrast, studies using the change score as the dependent variable found that participants with higher age benefit most from the training [41].

Table 4 Prognostic factors for training improvement in verbal long-term memory

Study	Test for outcome assessment	Dependent variable	Prognostic factor					
Multiple regression								
			Age	Education	Sex	Neuropsychology	Imaging	Others
O'Hara et al. [9]	Number of words correctly recalled.	Post-test scores Pre-test and change scores were integrated in regression.	↓	↑		Gain scores following training ↑ *		Length of training (short vs. long) ↑ Reported use of mnemonic at follow-up ↑ * Type of pre-training (standard vs. comprehensive) ↓ Pre-training ↑ *
Brooks et al. [8]	Proper name recall task	Post-test scores	↑ *			Pre-test score →*		Pre-training * Length → Length of training → Pre-training →
McDougall et al. [40]	RBMT	Change score Relative gains from beginning to end of training	↑	x				Ethnical group x
Park et al. [7]	Elderly verbal learning test, delayed recall *However, results are reported for "cognitive function" as outcome measure, which is not clearly defined*	Change score Post-pre	→	↓*	→	Pre-test scores of neuropsychological tests (Digit Span Test, Spatial Span Test, Categorical Fluency Test, short version of Boston Naming test) →		
Pesce et al. [20]	RVLT	Change score Post-pre						Change in dROMs ↓ Change in BAP ↑
Correlation analysis								
Leahy, Ridout, Mushtaq et al., [25]	Autobiographical memory specificity	Change score						Independence Depression Functional limitations Memory specificity
Andrewes et al. [26]	Laboratory Prospective Memory Assessment Everyday Prospective Memory Assessment	Change score				NART → Warrington Forced Choice Recognition → RAVT →		Mattis dementia scale → Depression →
Anschutz et al. [27]	Free recall of 2 lists Recognition of 2 lists	No clear reporting.	No clear reporting.					
Hill et al. [38]	Improvement in list recall	Change scores				MMSE ↑		
Leahy, Ridout, and Holland, [24]	Autobiographical memory specificity.	Change scores				Baseline cognitive flexibility ↑		
Group comparisons (ANOVA, *t* test)								
McDougall et al. [40]	RBMT	Pre-test and Post-test scores calculated in an ANOVA.	x	x				Ethnicity x

Table 4 Prognostic factors for training improvement in verbal long-term memory *(Continued)*

Study	Test for outcome assessment	Dependent variable	Prognostic factor
O'Hara et al., [43]	List-learning test	Pre-test and Post-test scores calculated in an ANOVA.	Apolipoprotein E4 ↓
Mixed models			
/			

Studies are sorted according to the statistical method used for obtaining the prognostic factors
ANOVA analysis of variance, *MMSE* Mini Mental State Examination, *NART* National Adult Reading Test, *RAVL* Rey Auditory Verbal Learning Test, *RBMT* Rivermead behavioural memory test, *RVLT* Rey Auditory Verbal Learning Test, *dROMs* reactive oxygen metabolites derivative compounds, *BAP* antioxidant levels; ↑ the higher the prognostic factor, the higher the improvement/positive correlation; ↓ the lower the prognostic factor, the higher the improvement/negative correlation; → no direction of effect reported; * significant; x unclear reporting

Table 5 Prognostic factors for training improvement in non-verbal short-term memory

Study	Test for outcome assessment	Dependent variable	Investigated prognostic factor					
Multiple regression								
			Age	Education	Sex	Neuropsychology	Imaging	Others
Park et al., [7]	Simple Rey Figure Test Immediate copy *However, results are reported for "cognitive function" as outcome measure, which is not clearly defined*	Change score Post-pre	→	↓*	→	Pre-test scores of neuropsychological tests (Digit Span Test, Spatial Span Test, Categorical Fluency Test, short version of Boston Naming test) →		
Correlation analysis								
Mohs et al. [21]	Biber Figure Learning Test	Post-test scores, Controlling for pre-test scores	→	→	→			Subjective reported memory →
Group comparisons (ANOVA, *t* test)								
/								
Mixed models								
/								

Studies are sorted according to the statistical method used for obtaining the prognostic factors. ↑ the higher the prognostic factor, the higher the improvement/positive correlation; ↓ the lower the prognostic factor, the higher the improvement/negative correlation; → no direction of effect reported; * significant; x unclear reporting

Table 6 Prognostic factors for training improvement in non-verbal long-term memory

Study	Test for outcome assessment	Dependent variable	Prognostic factor					
Multiple regression								
			Age	Education	Sex	Neuropsychology	Imaging	Others
Park et al. [7]	Simple Rey Figure Test Delayed Recall *However, results are reported for "cognitive function" as outcome measure, which is not clearly defined*	Change score Post-pre	→	↓*	→	Pre-test scores of neuropsychological tests (Digit Span Test, Spatial Span Test, Categorical Fluency Test, short version of Boston Naming test) →		
Correlation analysis								
Hampstead et al. [36]	Object Location Assignment accuracy	Modified change score Percentage of improvement relative to possible improvement after accounting for pre-test score				Trial Making Test B/A ↓ RBANS ↑	Amygdala volume ↑ Hippocampus volume ↑ Inferior lateral ventricles volume ↓	
Group comparisons (ANOVA, *t* test)								
McDougall et al. [40]	Brief Visuospatial Memory Test-Revised	ANOVA with pre- and post-test scores	→	→ *				Ethnicity—Hispanics and Blacks ↑* than Whites
O'Hara et al., [43]	Revised Benton Visual Retention Test	ANOVA with pre- and post-test scores						Apolipoprotein E4 ↓*
Mixed models								
/								

Studies are sorted according to the statistical method used for obtaining the prognostic factors. *RBANS* Repeatable Battery for the Assessment of Neuropsychological Status; ↑ the higher the prognostic factor, the higher the improvement/positive correlation; ↓ the lower the prognostic factor, the higher the improvement/negative correlation; → no direction of effect reported; * significant; x unclear reporting

Of the six studies that assessed *education* as a prognostic factor, it was shown that studies which used the post-test score as the dependent variable showed that participants with a higher educational level benefit most from the training [9, 42], whereas the study which used the change score as the dependent variable [7] again showed the opposite results indicating that participants with a lower educational level show improvements in their memory test performance. All other studies did not report data on the prognostic factor.

Sex was only investigated in two studies as a prognostic factor for changes in memory test performance [7, 21]. Yet, both studies did not provide any data on the direction of the prognostic factor.

Studies which used the post-test score as the dependent variable in their calculation to assess *neuropsychological test scores at study entry* showed that participants with higher neuropsychological test scores at study entry significantly benefited more from the memory training [42, 44, 45]. All other studies did not report any significant results on the prognostic factor.

Six studies investigated *brain imaging* prognostic factors. Two studies showed that when using standardized residuals as the dependent variable, a higher integrity of *white matter microstructure* was predictive for improvements in memory performance [32, 33]. Furthermore, two studies using the change score showed that a higher *hippocampal volume* was predictive for improvements in memory performance [29, 36]. Furthermore, a higher *activity in the frontal cortex* [22, 23] and higher *activity in the hippocampus* were predictive for changes in memory performance when using the change score as the dependent variable in the calculations.

Other investigated prognostic factors were *ethnicity*, *subjective reported memory*, *depression*, *openness to experience*, *extraversion*, *neuroticism*, *obesity*, *activities of daily living*, *apolipoprotein E4*, *length of training*, *biological antioxidant potential*, and *independence*. The only significant results of these prognostic factors were regarding *openness to experience*, showing that a higher value on the openness to experience scale predicted higher changes in memory test performance when using the change score as the dependent variable [35], and regarding *obesity*, showing that lower obesity scores predict improvements in memory performance when using the change score as the dependent variable [30, 31].

Discussion

This is the first systematic review that examines prognostic factors of changes in memory test performance after memory training in healthy older adults. The main findings are that (i) included studies used different types of dependent variables (change scores vs. post-test scores) when defining memory training success leading to contradictory results, and that (ii) age was the only variable investigated throughout most of the studies, showing that older adults showed improvements in memory test performance after training when using the change score as the dependent variable.

Methodological considerations

The most important result is that the direction of the relationship between the prognostic factor and the memory outcome (the more of x/ the less of x) differ depending on which dependent variable is evaluated as the outcome measure. For example, this means that studies that used post-test scores as the dependent outcome measure showed that participants with lower age showed greater improvements in memory test performance after training [9, 42, 44, 45] with only one exception [8]. However, it should be noted that the study of Brooks et al. [8] also integrated an interaction term in their analysis. In contrast, studies using the change score as the dependent variable found that participants with higher age benefit most from the training [41]. This finding is substantial for the interpretation of the reported findings in the current literature on prognostic factors of changes in memory test performance after memory training in healthy older adults: Until now, different directions of prognostic factors have been reported, but the cause of these differences have remained unresolved. Discussed explanations in single studies included characteristics of the used memory training, measurement procedures and the investigated sample [45, 46]. The present systematic review suggests, however, that these heterogeneous findings can mainly be explained by the different statistical methods used for prediction analyses so far, and the different dependent outcome measures (post-test scores vs. change scores vs. residual scores). Therefore, when reading and interpreting prognostic factor data of memory training improvement, our systematic review shows that it is of outstanding importance to take a closer look on the dependent variable used to measure training improvement.

Our systematic review shows that the included studies not only used different dependent variables but also different statistical methods to calculate prognostic factors (e.g., linear regression models, correlation analyses, mixed models, and group comparisons). However, not all used methods are suitable to answer the question of who benefits from memory training. For example, correlation analysis do not imply causal relationship and are therefore not an appropriate tool for measuring predictive performance as prognosis is defined as estimating the risk of future outcomes in individuals based on different characteristics. Also, group comparisons (e.g., *t* - tests, ANOVAs) are not suitable for prognostic factor measurement, because they only show group differences. Yet, there are no clear recommendations regarding the "proper way" to calculate

prognostic factors after memory training so far, even though it can be suggested that multiple regression analysis or structural equation models seem appropriate to answer the question of "who benefits" from training [47]. Smoleń et al. [47] suggest to use direct modeling of correlations between latent true measures and gain to investigate possible prognostic factors of changes in cognitive performances after CT.

Results of our review also show that investigated sample sizes in the included studies are often very small and that statistical power for the used calculations are lacking. It is important to note that the present review focuses on prognostic factors for memory performance after memory training instead of memory success after training.

Identified prognostic factors for changes in memory performance

The only prognostic factor that has been measured in several studies investigating verbal short- and long-term memory is "age." In studies which used the post-test score as the dependent variable [42, 44, 45], participants with younger age showed improvements after the memory training intervention, which may be explained by the magnification approach [48]. This account implies that participants who are already functioning at a high cognitive level can easily integrate new knowledge in already existing neuronal networks and can therefore profit faster and more easily from memory training. However, studies which use the change score as a dependent variable [41] show the opposite result: older participants benefited most from memory training. The latter result can be interpreted with the compensation hypothesis, stating that older participants may have more room for cognitive improvement [48]. This account implies that healthy older adults who are already functioning at optimal levels have less room for changes in memory training performance. When we look on the post-test performance, it is logical that younger participants who perform better at pretest also perform better after the training.

Further investigated prognostic factors include sociodemographic factors, neuropsychological test status at study entry in different domains, imaging measures, training characteristics, genetic variables (apolipoprotein E4), and personality traits. However, the reporting of most of the prognostic factors is insufficient so that only limited (or in some cases no) conclusions can be drawn from the data.

In one study, lower education was predictive for improvements in verbal long-term memory, non-verbal short-term memory, and non-verbal long-term memory when using the change score as a dependent variable [7]. These results might also be explained by the compensation hypothesis, showing that participants with less years of education show more room for cognitive improvement [48]. Yet, it is also important to keep in mind that the factor "education" might present more than just the years of schooling, but that it may be a proxy variable for socioeconomic status, early life factors, occupational health, or even the willingness to engage in lifelong learning or new activities [49–51]. All of these variables might affect the memory training performance and therefore additional variables should be taken into account in form of a prognostic model, to investigate the influence of years of education on training success while controlling for related covariates such as, e.g., socioeconomic status and cognitive reserve (which can be assessed with the help of questionnaires as the Lifetime of Experience Questionnaire [52]) or even also integrate these as possible further prognostic factors.

Regarding brain imaging factors, a higher hippocampal volume was a significant prognostic factor for improvements in memory performance after training in the domain verbal short-term memory [29]. However, it was not clearly reported which dependent variable was used in the study and therefore, clear conclusions of this result cannot be derived. In general, hippocampal-cortical connections are known to be critical for episodic memory functions [53], and it is known that the hippocampal volume is related to memory performance in older adults [54], and that memory training may enhance hippocampal activity [33]. Therefore, it seems plausible that a higher hippocampal volume constitutes a better "hardware" for memory plasticity. Further studies with a clear description and definition of the dependent variable used for measuring the prognostic effect of hippocampal volume on changes in memory test performance after memory training are needed to support this notion.

The apoE 4 allele, which is a well-known risk factor for Alzheimer disease [55] was a significant prognostic factor for improvements in memory test performance in non-verbal long-term memory. However, it was only assessed in a group comparison between carriers and non-carries of the allele, showing that non-carriers benefit more from training [43]. This finding is in line with a meta-analysis on the effects of apoE 4 on cognitive functions in non-impaired older adults [56], and a study on CT improvement of healthy older adults [46]. Interestingly, apoE and the apoE 4 human isoform both impair hippocampal neurogenesis and show therefore that apoE may influence hippocampal-related neurological diseases [57], showing a possible link between apoE 4 and hippocampal volume as prognostic factors of changes in memory test performance after memory training. However, further research is needed as only a limited number of

studies have investigated the effects of apoE 4 on training performance so far.

The one study that studied obesity as a possible prognostic factor for changes in memory test performance after memory training using the relative change score as the dependent variable [30, 31] found that older adults with obesity had a significantly lower training effect on the memory score than adults with normal weight. This result may be indicative for a relationship between obesity and impaired neural plasticity. There is evidence of an effect of obesity on inflammation, and onward an effect of inflammation on cognitive function [58]. Besides, there are several studies showing that obesity or high-fat feeding are associated with deficits in learning, memory, and executive functions [59, 60]. Due to the fact that the World Health Organization reports that the number of obese people (body mass index, BMI > 30) and overweight (BMI > 25) is reaching epidemic proportions worldwide [61], obesity is an important prognostic factor to further investigate.

Taken together, regarding sociodemographic factors (e.g., age, education), it seems that more "vulnerable" groups show stronger changes in memory test performance after memory training, while regarding biological factors (including the prognostic factors hippocampal volume, apoE 4, and obesity), the opposite pattern occurs—possibly meaning that the latter factors may serve as the "hardware" that functions as a driver of plasticity. However, evidence is far too rare to identify consistent patterns in order to formulate a clear hypothesis and more research is needed.

A further result of our systematic review is that throughout the studies, the choice of investigated prognostic factors is highly heterogeneous and seems often rather arbitrary than theory-based. This may be due to the fact that prognostic factor research is often a study "add-on" or a secondary or tertiary aim instead of the primary research question, and therefore constitutes an exploratory research approach. Yet, selective reporting of outcomes (and prognostic factors) is often a risk [62] and without pre-registration of studies, it is impossible to detect whether outcomes were assessed but not reported. Unfortunately, until now, pre-registration of prediction research is not mandatory [63].

Summarized, most of the prognostic factors reported in this systematic review are still highly under-investigated. In order to ensure an individual, personalized medicine approach, however, it is of high importance to identify special prognostic factors for changes in memory test performance after memory training to provide the best fitting nonpharmacological intervention approach for the individual's specific needs.

Reporting quality in the included studies

As already mentioned, the fact that prognostic factor calculation was often used as an "add-on" may contribute to several methodological short-comings in some studies. Therefore, this may also explain the overall poor reporting quality of the included studies. Especially prognostic factors and their statistical measures were not adequately described in most of the studies included in this review. This result is in line with other systematic reviews on prognostic factors in other research populations (e.g., participants with low back pain, participants with cancer) showing many methodological shortcomings in the design and conduct of studies that address prognosis [64, 65]. This shows that there is an immediate need for adequate reporting in the area of prognostic factors for changes in memory test performance after memory training—and more generally. The methodological shortcomings in the primary literature limit conclusions about prognostic factors for memory training success.

Limitations

When interpreting the results of this review, there are several limitations that have to be taken into account. First, it was difficult for the review authors to distinguish between prognostic factor and prognostic model studies, as the reporting was fairly poor in most studies. Most studies did not state whether their aim was to investigate a factor (the influence of one prognostic variable on the outcome), or a model (the influence of two or more prognostic variables and their interactions on the outcome). Further, the statistical methods were frequently not clearly reported so that in some cases, it was not possible to determine which prognostic variables were used in the final calculations. Therefore, a correct classification may not have been made in all included studies.

Furthermore, there was no scoring system regarding the assessment of the risk of bias tool QUIPS [19] to standardize the risk of bias assessment over other systematic reviews. However, a clear description of our risk of bias assessment procedure is provided in the Supplementary, so that traceability and replicability is provided.

In the present review, only studies published in English or German were included and therefore we may have missed studies published in other languages. As a further limitation, the present systematic review only focuses on memory outcomes after memory training, hereby disregarding other cognitive domains, as well as other non-cognitive outcomes (e.g., depression, quality of life, activities of daily living), and other single-domain (e.g., working memory training) and multi-domain CT, respectively. Further systematic reviews are needed to elaborate the knowledge on prognostic factors of CT success.

Unfortunately, we could not perform a meta-analysis on the investigated prognostic factors of memory training success as planned and described in the preregistration of this systematic review (ID: CRD42019127479, https://www.crd.york.ac.uk/PROSPERO/). This had mainly two reasons: First, in most of the studies not enough or no statistical data at all was provided on the investigated prognostic factors, and second, the overall statistical reporting was too poor to extract the necessary details. Furthermore, due to the use of the different dependent variables, we could not integrate all available data in one single analysis without falsifying the results. When trying to calculate different analyses for the different dependent variables, we then had not enough data again to conduct the analyses.

Strengths of this systematic review

A particular strength of the study is that it is the first review that focuses on prognostic factors for changes in memory test performance after memory training in healthy older adults. This systematic work was able to shed light on the reasons of inconsistent results of research regarding prognostic factors in the literature: they seem to be mainly due to different used methodological approaches.

A further strength is that the present review was conducted using Cochrane standards for systematic reviews. The present review further provides a differentiation among the different memory outcomes and a detailed reporting of the statistical methods of the included studies.

Implications for further prognostic research

Yet, the results and conclusions regarding the statistical analysis of the prognostic factors for changes in memory test performance after memory training might also be transferred to other trainings and cognitive outcomes. As a clear recommendation, independent of the investigated non-pharmacological intervention and the investigated outcome, one should be aware of the used dependent variable and statistical methods to assess prognostic factors. We recommend the use of the change score as a dependent variable to answer the question "who benefits" from a nonpharmacological intervention and to use multiple regression analysis or structural equation models instead of correlation analysis and group comparisons.

Conclusion

This present systematic review on prognostic factors of changes in memory test performance after memory training shows substantial short-comings in methodological reporting and statistical analyses and emphasizes the need of elaborated prognostic factor studies with large sample sizes, clear descriptions of prognostic factor and confounder measurement, and clear reporting standards. Furthermore, a special focus should clearly be on the use of the dependent variables used for prognostic factor calculation. Our systematic review also showed that most prognostic factors are still highly underinvestigated. Prognostic factor research should not be an "add-on" to already existing studies, but should be a separate focus following clear reporting and conduction guidelines, as prognostic factor research is of high importance for aiding treatment and lifestyle decisions, improving individual dementia risk prediction, and providing new treatment options [6]. As a preliminary conclusion, regarding prognostic factors for changes in memory test performance after memory training, older adults seem to show greater improvements in memory test performance after memory training than younger adults.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s41512-020-0071-8 .

Additional file 1. Table 1. The PRISMA for Abstracts Checklist. Table 2. The PRISMA checklist for systematic reviews. Table 3. Prognostic models for memory training success in healthy older adults, search strategy (CENTRAL). Table 4. Prognostic models for memory training success in healthy older adults, search strategy (Medline). Table 5. Prognostic models for memory training success in healthy older adults, search strategy (PsycInfo). Table 6. Prognostic models for memory training success in healthy older adults, search strategy (Web of Science Core Collection). Table 7. Risk of Bias Assessment using the QUIPS tool. Table 8. Outcomes, prognostic factors and details on analysis of the included studies. Note. Abbreviations: RAVLT: Rey Auditory Verbal Learning Task; BDI: Beck Depression Inventory; NART: National Adult Reading Test; ALFF: Amplitude of low-frequency fluctuation; fALFF: Fractional amplitude of low-frequency fluctuation; BMI: body mass index; MMSE: Mini-Mental Status Examination; RBANS: Repeatable Battery for the Assessment of Neuropsychological Status; HVLT-R: Hopkins Verbal Learning Test-Revised; AMT: Autobiographical Memory Task; RBMT: Rivermead Behavioral Memory Test; BVMT-R: Brief Visuospatial Memory Test revised; EMS: Elderly Memory Disorder Scale; BVRT: Revised Benton Visual Retention Test; MEPS: Means End Problem Solving Procedure; FLP: functional limitation profile; FU: Follow-up; HADS: Hospital Anxiety and Depression Scale; IADL: Instrumental and basic activities of daily living; NEO-PI:NEO Personality Inventory; ZSRDS: Zung Self-Rating Depression Scale; ACE-III: Addenbrooke's Cognitive Examination-III. Table 9. Overview of study results. Abbreviations: AMT: Autobiographical Memory Task; BVRT: Revised Benton Visual Retention Test; MMSE: Mini-Mental Status Examination. NEO-PI:NEO Personality Inventory, MEPS: Means End Problem Solving Procedure; FLP: functional limitation.

Acknowledgements

Not applicable.

Authors' contributions

MR and EK conceived the presented idea. MR and AKF conducted the systematic search, extracted the data, and conducted the quality assessment with the help of FK. NS contributed to the systematic search and data extraction. MR took the lead in writing the manuscript. All authors provided critical feedback and helped shape the research and manuscript. EK supervised the project. All authors read and and approved the final manuscript

Funding
The study was funded by the Brandau-Laibach Stiftung, Cologne.

Availability of data and materials
The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
EK has received grants from the German Ministry of Education and Research, ParkinsonFonds Deutschland GmbH, the German Parkinson Society; honoraria from: Oticon GmbH, Hamburg, Germany; Lilly Pharma GmbH, Bad Homburg, Germany; Bernafon AG, Bern, Switzerland; Desitin GmbH, Hamburg, Germany. EK is author of the cognitive training program NEUROvitalis but receives no corresponding honoraria. AKF has received a grant from the German Parkinson Society, and honoraria from ProLog Wissen GmbH, Cologne, Germany and pro audito Switzerland, Zürich, Switzerland. AFK is author of the cognitive training program NEUROvitalis but receives no corresponding honoraria. MR has received a grant from the Brandau-Laibach Stiftung, and a grant from the German Ministry of Education and Research. FK and NS do not declare any conflict of interests.

Author details
[1]Department of Medical Psychology | Neuropsychology and Gender Studies & Center for Neuropsychological Diagnostics and Intervention (CeNDI), Faculty of Medicine and University Hospital Cologne, University of Cologne, Kerpener Str. 68, 50937 Cologne, Germany. [2]Evidence-Based Oncology, Department I of Internal Medicine, Faculty of Medicine and University Hospital Cologne, University of Cologne, Kerpener Str. 62, 50937 Cologne, Germany.

Received: 8 October 2019 Accepted: 13 February 2020
Published online: 21 May 2020

References

1. Reuter-Lorenz PA, Park DC. How does it STAC up? Revisiting the scaffolding theory of aging and cognition. Neuropsychol Rev. 2014;24(3):355–70 https://doi.org/10.1007/s11065-014-9270-9.
2. Kallio E-L, Öhman H, Kautiainen H, Hietanen M, Pitkälä K. Cognitive training interventions for patients with Alzheimer's disease: a systematic review. Journal of Alzheimer's Disease : JAD. 2017;56(4):1349–72 https://doi.org/10.3233/JAD-160810.
3. Chiu H-L, Chu H, Tsai J-C, Liu D, Chen Y-R, Yang H-L, Chou K-R. The effect of cognitive-based training for the healthy older people: A meta-analysis of randomized controlled trials. PloS One. 2017;12(5):e0176742 https://doi.org/10.1371/journal.pone.0176742.
4. Lampit A, Hallock H, Valenzuela M. Computerized cognitive training in cognitively healthy older adults: a systematic review and meta-analysis of effect modifiers. PLoS Med. 2014;11(11):e1001756 https://doi.org/10.1371/journal.pmed.1001756.
5. Salthouse TA. Effects of age and ability on components of cognitive change. Intelligence. 2013;41(5):501–11 https://doi.org/10.1016/j.intell.2013.07.005.
6. Riley RD, Hayden JA, Steyerberg EW, Moons KGM, Abrams K, Kyzas PA, et al. Prognosis research strategy (PROGRESS) 2: Prognostic factor research. PLoS Med. 2013;10(2):e1001380 https://doi.org/10.1371/journal.pmed.1001380.
7. Park S, Ryu S-H, Yoo Y, Yang J-J, Kwon H, Youn J-H, et al. Neural predictors of cognitive improvement by multi-strategic memory training based on metamemory in older adults with subjective memory complaints. Sci Rep. 2018;8(1):1095 https://doi.org/10.1038/s41598-018-19390-2.
8. Brooks JO, Friedman L, Pearman AM, Gray C, Yesavage JA. Mnemonic training in older adults: Effects of age, length of training, and type of cognitive pretraining. Int Psychogeriatr. 1999;11(1):75–84.
9. O'Hara R, Brooks JO, Friedman L, Schröder CM, Morgan KS, Kraemer HC. Long-term effects of mnemonic training in community-dwelling older adults. J Psychiatric Res. 2007;41(7):585–90 https://doi.org/10.1016/j.jpsychires.2006.04.010.
10. Verhaeghen P, Marcoen A, Goossens L. Improving memory performance in the aged through mnemonic training: A meta-analytic study. Psychol Aging. 1992;7(2):242–51.
11. Fairchild JK, Friedman L, Rosen AC, Yesavage JA. Which older adults maintain benefit from cognitive training? Use of signal detection methods to identify long-term treatment gains. Int Psychogeriatr. 2013;25(4):607–16 https://doi.org/10.1017/S1041610212002049.
12. Riley RD, Moons KGM, Snell KIE, Ensor J, Hooft L, Altman DG, et al. A guide to systematic review and meta-analysis of prognostic factor studies. BMJ. 2019;364:k4597 https://doi.org/10.1136/bmj.k4597.
13. Riley RD, Ridley G, Williams K, Altman DG, Hayden J, de Vet HCW. Prognosis research: Toward evidence-based results and a Cochrane methods group. J Clin Epidemiol. 2007;60(8):863–5 author reply 865-6. https://doi.org/10.1016/j.jclinepi.2007.02.004.
14. Debray TPA, Damen JAAG, Snell KIE, Ensor J, Hooft L, Reitsma JB, et al. A guide to systematic review and meta-analysis of prediction model performance. BMJ. 2017;356:i6460 https://doi.org/10.1136/bmj.i6460.
15. Moons KGM, de Groot JAH, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: The CHARMS checklist. PLoS Med. 2014;11(10):e1001744 https://doi.org/10.1371/journal.pmed.1001744.
16. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLoS Med. 2009;6(7):e1000097 https://doi.org/10.1371/journal.pmed.1000097.
17. Veritas Health Innovation. Covidence systematic review software. Melbourne; n.d. Retrieved from www.covidence.org..
18. Cowan N. Chapter 20 What are the differences between long-term, short-term, and working memory? In: Sossin WS, editor. Progress in Brain Research: Vol. 169. Progress in brain research: Essence of memory. Amsterdam: Elsevier; 2008. p. 323–38. https://doi.org/10.1016/S0079-6123(07)00020-9.
19. Hayden JA, van der Windt DA, Cartwright JL, Côté P, Bombardier C. Assessing bias in studies of prognostic factors. Ann Int Med. 2013;158(4):280–6 https://doi.org/10.7326/0003-4819-158-4-201302190-00009.
20. Pesce M, Tatangelo R, La Fratta I, Rizzuto A, Campagna G, Turli C, et al. Aging-related oxidative stress: positive effect of memory training. Neuroscience. 2018;370:246–55 https://doi.org/10.1016/j.neuroscience.2017.09.046.
21. Mohs RC, Ashman TA, Jantzen K, Albert M, Brandt J, Gordon B, et al. A study of the efficacy of a comprehensive memory enhancement program in healthy elderly persons. Psychiatr Res. 1998;77(3):183–95.
22. Kirchhoff BA, Anderson BA, Smith SE, Barch DM, Jacoby LL. Cognitive training-related changes in hippocampal activity associated with recollection in older adults. Neuro Image. 2012b;62(3):1956–64 https://doi.org/10.1016/j.neuroimage.2012.06.017.
23. Kirchhoff BA, Anderson BA, Jacoby LL. Cognitive and neural effects of semantic encoding strategy training in older adults. Cereb Cortex. 2012a;22(4):788–99 https://doi.org/10.1093/cercor/bhr129.
24. Leahy F, Ridout N, Holland C. Memory flexibility training for autobiographical memory as an intervention for maintaining social and mental well-being in older adults. Memory. 2018a;26(9):1310–22 https://doi.org/10.1080/09658211.2018.1464582.
25. Leahy F, Ridout N, Mushtaq F, Holland C. Improving specific autobiographical memory in older adults: impacts on mood, social problem solving, and functional limitations. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn. 2018b;25(5):695–723 https://doi.org/10.1080/13825585.2017.1365815.
26. Andrewes DG, Kinsella G, Murphy M. Using a memory handbook to improve everyday memory in community-dwelling older adults with memory complaints. Exp Aging Res. 1996;22(3):305–22 https://doi.org/10.1080/03610739608254013.
27. Anschutz L, Camp CJ, Markley RP, Kramer JJ. Remembering mnemonics: a three-year follow-up on the effects of mnemonics training in elderly adults. Exp Aging Res. 1987;13(3):141–3 https://doi.org/10.1080/03610738708259315.
28. Bissig D, Lustig C. Who benefits from memory training? Psychol Sci. 2007;18(8):720–6 https://doi.org/10.1111/j.1467-9280.2007.01966.x.
29. Bråthen ACS, de Lange A-MG, Rohani DA, Sneve MH, Fjell AM, Walhovd KB. Multimodal cortical and hippocampal prediction of episodic-memory

plasticity in young and older adults. Hum Brain Mapp. 2018;39(11):4480–92 https://doi.org/10.1002/hbm.24287.
30. Clark DO, Xu H, Callahan CM, Unverzagt FW. Does body mass index modify memory, reasoning, and speed of processing training effects in older adults. Obesity (Silver Spring). 2016a;24(11):2319–26 https://doi.org/10.1002/oby.21631.
31. Clark DO, Xu H, Unverzagt FW, Hendrie H. Does targeted cognitive training reduce educational disparities in cognitive function among cognitively normal older adults? Int J Geriatric Psychiatr. 2016b;31(7):809–17 https://doi.org/10.1002/gps.4395.
32. de Lange A-MG, Bråthen ACS, Grydeland H, Sexton C, Johansen-Berg H, Andersson JLR, et al. White matter integrity as a marker for cognitive plasticity in aging. Neurobiol Aging. 2016;47:74–82 https://doi.org/10.1016/j.neurobiolaging.2016.07.007.
33. de Lange A-MG, Bråthen ACS, Rohani DA, Grydeland H, Fjell AM, Walhovd KB. The effects of memory training on behavioral and microstructural plasticity in young and older adults. Hum Brain Mapp. 2017;38(11):5666–80 https://doi.org/10.1002/hbm.23756.
34. Tomaszewski Farias S, Giovannetti T, Payne BR, Marsiske M, Rebok GW, Schaie KW, et al. Self-perceived difficulties in everyday function precede cognitive decline among older adults in the ACTIVE study. J Int Neuropsychol Soc. 2018;24(1):104–12 https://doi.org/10.1017/S1355617717000546.
35. Finkel SI, Yesavage JA. Learning mnemonics: A preliminary evaluation of a computer-aided instruction package for the elderly. Exp Aging Res. 1989;15(3-4):199–201 https://doi.org/10.1080/03610738908259776.
36. Hampstead BM, Sathian K, Phillips PA, Amaraneni A, Delaune WR, Stringer AY. Mnemonic strategy training improves memory for object location associations in both healthy elderly and patients with amnestic mild cognitive impairment: A randomized, single-blind study. Neuropsychology. 2012;26(3):385–99 https://doi.org/10.1037/a0027545.
37. Hill RD, Sheikh JI, Yesavage J. The effect of mnemonic training on perceived recall confidence in the elderly. Exp Aging Res. 1987;13(4):185–8 https://doi.org/10.1080/03610738708259323.
38. Hill RD, Yesavage JA, Sheikh J, Friedman L. Mental status as a predictor of response to memory training in older adults. Educ Gerontol. 1989;1989:633–8.
39. López-Higes R, Rodríguez-Rojo IC, Prados JM, Montejo P, Del-Río D, Delgado-Losada ML, et al. Apoe ε4 modulation of training outcomes in several cognitive domains in a sample of cognitively intact older adults. J Alzheimers Dis. 2017;58(4):1201–15 https://doi.org/10.3233/JAD-161014.
40. McDougall GJ, Becker H, Pituch K, Acee TW, Vaughan PW, Delville CL. Differential benefits of memory training for minority older adults in the SeniorWISE study. Gerontologist. 2010a;50(5):632–45 https://doi.org/10.1093/geront/gnq017.
41. McDougall GJ, Becker H, Pituch K, Acee TW, Vaughan PW, Delville CL. The SeniorWISE study: Improving everyday memory in older adults. Arch Psychiatr Nurs. 2010b;24(5):291–306 https://doi.org/10.1016/j.apnu.2009.11.001.
42. Neely AS, Bäckman L. Effects of multifactorial memory training in old age: Generalizability across tasks and individuals. J Gerontol B Psychol Sci Soc Sci. 1995;50(3):134–40.
43. O'Hara R, Yesavage JA, Kraemer HC, Mauricio M, Friedman LF, Murphy GM. The APOE epsilon4 allele is associated with decline on delayed recall performance in community-dwelling older adults. J Am Geriatr Soc. 1998;46(12):1493–8.
44. Rosi A, Del Signore F, Canelli E, Allegri N, Bottiroli S, Vecchi T, Cavallini E. The effect of strategic memory training in older adults: Who benefits most? Int Psychogeriatr. 2018;30(8):1235–42 https://doi.org/10.1017/S1041610217002691.
45. Sandberg P, Rönnlund M, Derwinger-Hallberg A, Stigsdotter Neely A. Memory plasticity in older adults: Cognitive predictors of training response and maintenance following learning of number-consonant mnemonic. Neuropsychol Rehabil. 2016;26(5-6):742–60 https://doi.org/10.1080/09602011.2015.1046459.
46. Roheger M, Meyer J, Kessler J, Kalbe E. Predicting short- and long-term cognitive training success in healthy older adults: Who benefits? Neuropsychol Dev Cogn B Aging Neuropsychol Cogn. 2019:1–19 https://doi.org/10.1080/13825585.2019.1617396.
47. Smoleń T, Jastrzebski J, Estrada E, Chuderski A. Most evidence for the compensation account of cognitive training is unreliable. Memory Cogn. 2018;46(8):1315–30 https://doi.org/10.3758/s13421-018-0839-z.
48. Lövdén M, Brehmer Y, Li S-C, Lindenberger U. Training-induced compensation versus magnification of individual differences in memory performance. Front Hum Neurosci. 2012;6:141 https://doi.org/10.3389/fnhum.2012.00141.
49. Krieger N, Williams DR, Moss NE. Measuring social class in US public health research: Concepts, methodologies, and guidelines. Annu Rev Public Health. 1997;18:341–78 https://doi.org/10.1146/annurev.publhealth.18.1.341.
50. Langbaum JBS, Rebok GW, Bandeen-Roche K, Carlson MC. Predicting memory training response patterns: results from ACTIVE. J Gerontol B Psychol Sci Soc Sci. 2009;64(1):14–23 https://doi.org/10.1093/geronb/gbn026.
51. Leigh JP, Fries JF. Education, gender, and the compression of morbidity. Int J Aging Hum Dev. 1994;39(3):233–46 https://doi.org/10.2190/XQXR-UTGP-WA8X-9FQJ.
52. Valenzuela, M. J., & Sachdev, P. Assessment of complex mental activity across the lifespan: development of the Lifetime of Experiences Questionnaire (LEQ). Psychological medicine. 2007;37(7):1015–25.
53. Buzsáki G, Moser EI. Memory, navigation and theta rhythm in the hippocampal-entorhinal system. Nat Neurosci. 2013;16(2):130–8 https://doi.org/10.1038/nn.3304.
54. Pohlack ST, Meyer P, Cacciaglia R, Liebscher C, Ridder S, Flor H. Bigger is better! Hippocampal volume and declarative memory performance in healthy young men. Brain Struc Funct. 2014;219(1):255–67 https://doi.org/10.1007/s00429-012-0497-z.
55. Liu C-C, Liu C-C, Kanekiyo T, Xu H, Bu G. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. Nat Rev Neurol. 2013;9(2):106–18 https://doi.org/10.1038/nrneurol.2012.263.
56. Wisdom NM, Callahan JL, Hawkins KA. The effects of apolipoprotein E on non-impaired cognitive functioning: A meta-analysis. Neurobiol Aging. 2011;32(1):63–74 https://doi.org/10.1016/j.neurobiolaging.2009.02.003.
57. Tensaouti Y, Stephanz EP, Yu T-S, Kernie SG. Apoe regulates the development of adult newborn hippocampal neurons. ENeuro. 2018;5(4) https://doi.org/10.1523/ENEURO.0155-18.2018.
58. Miller AA, Spencer SJ. Obesity and neuroinflammation: a pathway to cognitive impairment. Brain Behav Immun. 2014;42:10–21 https://doi.org/10.1016/j.bbi.2014.04.001.
59. Elias MF, Elias PK, Sullivan LM, Wolf PA, D'Agostino RB. Lower cognitive function in the presence of obesity and hypertension: The Framingham heart study. Int J Obes Relat Metab Disord. 2003;27(2):260–8 https://doi.org/10.1038/sj.ijo.802225.
60. Sabia S, Kivimaki M, Shipley MJ, Marmot MG, Singh-Manoux A. Body mass index over the adult life course and cognition in late midlife: The Whitehall II Cohort Study. Am J Clin Nutr. 2009;89(2):601–7 https://doi.org/10.3945/ajcn.2008.26482.
61. World Health Organization (2013). Obesity and overweight. Retrieved from https://www.who.int/.
62. Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, Smyth R, Williamson PR. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. BMJ. 2010;340:c365 https://doi.org/10.1136/bmj.c365.
63. Bouwmeester W, Zuithoff NPA, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, et al. Reporting and methods in clinical prediction research: A systematic review. PLoS Med. 2012;9(5):1–12 https://doi.org/10.1371/journal.pmed.1001221.
64. Hayden JA, Chou R, Hogg-Johnson S, Bombardier C. Systematic reviews of low back pain prognosis had variable methods and results: Guidance for future prognosis reviews. J Clin Epidemiol. 2009;62(8):781–796.e1 https://doi.org/10.1016/j.jclinepi.2008.09.004.
65. Riley RD, Sauerbrei W, Altman DG. Prognostic markers in cancer: The evolution of evidence from single studies to meta-analysis, and beyond. Brit J Cancer. 2009;100(8):1219–29 https://doi.org/10.1038/sj.bjc.6604999.
66. Veritas Health Innovation. Covidence systematic review software. Melbourne; n.d. Retrieved from www.covidence.org.

Publisher's Note

A Systematic Review on Predictors of Working Memory Training Responsiveness in Healthy Older Adults: Methodological Challenges and Future Directions

Anja Ophey[1*], Mandy Roheger[1], Ann-Kristin Folkerts[1], Nicole Skoetz[2], Elke Kalbe, Prof. Dr.[1*]

[1]Department of Medical Psychology | Neuropsychology & Gender Studies, University Hospital of Cologne, Germany, [2]Center of Integrated Oncology Aachen Bonn Cologne Duesseldorf, Department I of Internal Medicine, University Hospital of Cologne, Germany

Submitted to Journal:
Frontiers in Aging Neuroscience

Article type:
Systematic Review Article

Manuscript ID:
575804

Received on:
24 Jun 2020

Revised on:
25 Aug 2020

Frontiers website link:
www.frontiersin.org

Conflict of interest statement

The authors declare a potential conflict of interest and state it below

AO reports no conflicts of interest.
MR has received a grant from the Brandau-Laibach Stiftung, and a grant from the German Ministry of Education and Research.
AKF has received a grant from the German Parkinson Society, and honoraria from ProLog Wissen GmbH, Cologne, Germany, pro audito Switzerland, Zürich, Switzerland, Seminar- und Fortbildungszentrum Rheine, Germany, and LOGOMANIA, Fendt & Sax GbR, Munich, Germany. AFK is author of the cognitive training program NEUROvitalis but receives no corresponding honoraria.
NS reports no conflicts of interest.
EK has received grants from the German Ministry of Education and Research, ParkinsonFonds Deutschland gGmbH, the German Parkinson Society; honoraria from: Oticon GmbH, Hamburg, Germany; Lilly Pharma GmbH, Bad Homburg, Germany; Bernafon AG, Bern, Switzerland; Desitin GmbH, Hamburg, Germany. EK is author of the cognitive training program NEUROvitalis but receives no corresponding honoraria.

Author contribution statement

AO, MR, and EK conceptualized the presented work. MR conducted the systematic search, NS contributed to the systematic search. AO, MR, and AKF conducted the title and abstract screening. AO and MR conducted the full text screening, extracted the data, and conducted the risk of bias assessment. AO drafted the first version of the manuscript. All authors revised the manuscript for intellectual content and approved the final version of the manuscript. EK supervised the project during each stage of work.

Keywords

prognostic review, Systematic review, healthy aging, working memory training, training responsiveness, individual differences

Abstract

Word count: 350

Background: Research on predictors of working memory training responsiveness, which could help to tailor cognitive interventions individually, are a timely topic in healthy aging. However, findings are highly heterogeneous, reporting partly conflicting results following a broad spectrum of methodological approaches to answer the question “who benefits most” from working memory training.
Objective: The present systematic review aimed to systematically investigate prognostic factors and models for working memory training responsiveness in healthy older adults.
Method: Four online data bases were searched up to October 2019 (MEDLINE Ovid, Web of Science, CENTRAL, PsycINFO). Inclusion criteria for full-texts were publication in a peer-reviewed journal in English/German, inclusion of healthy older individuals aged ≥55 years without any neurological and/or psychiatric diseases including cognitive impairment, and the investigation of prognostic factors and/or models for training responsiveness after targeted working memory training in terms of direct training effects, near-transfer effects to verbal and visuospatial working memory, as well as far-transfer effects to other cognitive domains and behavioural variables. Study design was not limited to randomized controlled trials.
Results:16 studies including n=675 healthy older individuals with a mean age of 63.0-86.8 years were included in this review. Within these studies, 5 prognostic model approaches and 18 factor finding approaches were reported. Risk of bias was assessed using the Quality-in-Prognosis-Studies-checklist, indicating that important information, especially regarding the domains study attrition, study confounding and statistical analysis, and reporting, was lacking throughout many of the investigated studies. Age, education, intelligence, and baseline performance in working memory or other cognitive domains were frequently investigated predictors across studies.
Conclusions: Given the methodological shortcomings of the included studies, no clear conclusions can be drawn, and emerging patterns of prognostic effects will have to survive sound methodological replication in future attempts to promote precision medicine approaches in the context of working memory training. Methodological considerations are discussed and our findings are embedded to the cognitive aging literature, considering for example the cognitive reserve framework and the compensation versus magnification account. The need for personalized cognitive prevention and intervention methods to counteract cognitive decline in the aging population is high and the potential enormous.
Registration:PROSPERO,CRD42019142750

Contribution to the field

One key aspect of healthy aging is the maintenance of cognitive functions by preventing or delaying the onset of cognitive dysfunction. In this context, working memory has become a main target for cognitive training interventions. However, there is an ongoing debate on the effectiveness of targeted working memory training (WMT). Given those heterogeneous results, identifying modifying, otherwise called “prognostic” or “moderating” factors (including both individual- and training-related characteristics) of WMT responsiveness on a single-study level is a growing field, however, no consensus regarding the question “who benefits most from WMT” was reached yet. Therefore, the present systematic review aims to systematically investigate prognostic factors and models for WMT responsiveness in healthy older adults. Summarizing, a pattern emerged in which individuals with younger age, less education, lower baseline performance, and higher intelligence benefit most from working memory training. Our findings are discussed under methodological considerations and embedded to the cognitive aging literature, considering for example the cognitive reserve framework and the compensation versus magnification account. By taking into account individual differences in cognitive plasticity and following responsiveness to cognitive training interventions, our findings contribute to the need for personalized cognitive prevention and intervention methods to counteract cognitive decline in the aging population.

Funding statement

This publication was funded by the Brandau-Laibach-Stiftung. No additional funding by funding agencies in the public, commercial, or not-for-profit sectors was obtained.

Data availability statement

Generated Statement: The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

A Systematic Review on Predictors of Working Memory Training Responsiveness in Healthy Older Adults: Methodological Challenges and Future Directions

Anja Ophey[1], Mandy Roheger[1*], Ann-Kristin Folkerts[1], Nicole Skoetz[2], Elke Kalbe[1*]

[1] Department of Medical Psychology | Neuropsychology & Gender Studies, Center for Neuropsychological Diagnostics and Intervention (CeNDI), Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany

[2] Center for Integrated Oncology Aachen Bonn Cologne Duesseldorf, Department I of Internal Medicine, Faculty of Medicine and University Hospital of Cologne, University of Cologne, Cologne, Germany

*** Correspondence:**
Prof. Dr. Elke Kalbe, Kerpener Str. 62, 50937 Cologne, Germany, *phone* +49 221 478-6669, *fax* +49 221 478-3420, elke.kalbe@uk-koeln.de

Words: 9560

Number of Tables: 3

Number of Figures: 1

Supplementary Material

Keywords: Prognostic Review, Systematic Review, Healthy Aging, Working Memory Training, Training Responsiveness.

Abstract

Background: Research on predictors of working memory training responsiveness, which could help to tailor cognitive interventions individually, are a timely topic in healthy aging. However, findings are highly heterogeneous, reporting partly conflicting results following a broad spectrum of methodological approaches to answer the question "who benefits most" from working memory training.

Objective: The present systematic review aimed to systematically investigate prognostic factors and models for working memory training responsiveness in healthy older adults.

Method: Four online data bases were searched up to October 2019 (MEDLINE Ovid, Web of Science, CENTRAL, PsycINFO). Inclusion criteria for full-texts were publication in a peer-reviewed journal in English/German, inclusion of healthy older individuals aged ≥55 years without any neurological and/or psychiatric diseases including cognitive impairment, and the investigation of prognostic factors and/or models for training responsiveness after targeted working memory training in terms of direct training effects, near-transfer effects to verbal and visuospatial working memory, as well as far-transfer effects to other cognitive domains and behavioural variables. Study design was not limited to randomized controlled trials.

Results: 16 studies including *n*=675 healthy older individuals with a mean age of 63.0-86.8 years were included in this review. Within these studies, 5 prognostic model approaches and 18 factor finding approaches were reported. Risk of bias was assessed using the Quality-in-Prognosis-Studies-checklist, indicating that important information, especially regarding the domains study attrition, study confounding and statistical analysis, and reporting, was lacking throughout many of the investigated studies. Age, education, intelligence, and baseline performance in working memory or other cognitive domains were frequently investigated predictors across studies.

Conclusions: Given the methodological shortcomings of the included studies, no clear conclusions can be drawn, and emerging patterns of prognostic effects will have to survive sound methodological replication in future attempts to promote precision medicine approaches in the context of working memory training. Methodological considerations are discussed and our findings are embedded to the cognitive aging literature, considering for example the cognitive reserve framework and the compensation versus magnification account. The need for personalized cognitive prevention and intervention methods to counteract cognitive decline in the aging population is high and the potential enormous.

Registration: PROSPERO, ID CRD42019142750

1 Introduction

The promotion of healthy aging constitutes a major goal given the demographic change that the world's population is facing (Parish et al., 2019). One key aspect of healthy aging is the maintenance of cognitive functions by preventing or delaying the onset of clinically relevant cognitive dysfunction or even reversing age-related cognitive decline (Lustig, Shah, Seidler, & Reuter-Lorenz, 2009). Cognitive decline is one of the most feared aspects in aging (Deary et al., 2009), as it reduces the quality of life of both the aging individual and his/her relatives and increases the burden on care providers and the public healthcare system. A decline of executive functions, working memory, processing speed, and memory – cognitive functions that are essential for everyday functioning – are the most prominent cognitive alterations in healthy aging (Paraskevoudi, Balcı, & Vatakis, 2018). Especially working memory, a capacity-limited system for short-term storage and manipulation of information, is of fundamental importance for general cognitive functioning and is seen as a key function and processing resource for other cognitive abilities (Chai, Abd Hamid, & Abdullah, 2018; Salthouse, 1990).

Cognitive training interventions as a non-pharmacological intervention and prevention method have gained increased scientific interest (Lustig et al., 2009). A recent meta-analysis of Chiu et al. (2017) on broad cognitive interventions in healthy older adults clearly indicated the potential of cognitive interventions to counteract cognitive decline. However, some issues such as the degree of transfer to untrained tasks and long-term effects remain a matter of debate. In this context, working memory has become a main target for cognitive training interventions. The role of working memory as a processing resource for other cognitive abilities (Chai et al., 2018; Salthouse, 1990) implies that working memory improvements after targeted working memory training (WMT) might naturally lead to positive transfer effects to other cognitive functions and even fluid intelligence (Au et al., 2015). Despite a general consensus regarding the effectiveness of targeted WMT regarding direct training effects (i.e. effects in trained working memory tasks over the course of training) and near-transfer effects (i.e. effects in untrained working memory tasks), evidence on far-transfer effects (i.e. effects in untrained domains) for different populations including healthy older adults has not convincingly been shown (for recent meta-analyses see e.g. Melby-Lervåg, Redick, & Hulme, 2016; Sala, Aksayli, Tatlidil, Gondo, & Gobet, 2019; Soveri, Antfolk, Karlsson, Salo, & Laine, 2017; Teixeira-Santos et al., 2019; Weicker, Villringer, & Thöne-Otto, 2016). Given those heterogeneous results concerning effects after WMT, identifying modifying, so-called "prognostic" or "moderating" factors (including both individual- and training-related characteristics) of WMT responsiveness seems highly relevant.

In general, a prognostic factor is defined as any measure that, among people with a given condition (e.g. the process of aging), is associated with a subsequent outcome (e.g. changes in cognition after certain interventions) (Riley et al., 2013). In prognostic research, prognostic factor finding studies and prognostic model studies are distinguished: Prognostic factor finding studies aim at establishing one or several variables as independent prognostic factors associated with an outcome. In contrast, prognostic model studies identify more than one prognostic factor, assign relative weights to each prognostic factor, and estimate the model's predictive performance through calibration and discrimination (Moons, Royston, Vergouwe, Grobbee, & Altman, 2009). Identifying prognostic factors for individual treatment response to WMT would take into account individual differences in cognitive plasticity and following responsiveness to cognitive training interventions (Baltes & Lindenberger, 1988; Bürki, Ludwig, Chicherio, & de Ribaupierre, 2014; Noack, Lövdén, Schmiedek, & Lindenberger, 2009). It would further contribute to the development of an encompassing approach in terms of a "personalized" or "precision medicine" (Hingorani et al., 2013)

in healthy aging and the prevention of cognitive decline, for example in the context of Alzheimer's disease (Berkowitz et al., 2018; Reitz, 2016).

The latest meta-analysis on WMT for healthy older adults (Teixeira-Santos et al., 2019) included a broad moderator analysis for WMT responsiveness. Despite training-related variables (e.g. training dose and length, number of sessions, training type), study population characteristics (e.g. age, education, general cognitive ability, baseline performance) were considered as moderating variables (Teixeira-Santos et al., 2019). The meta-analysis mainly identified training-related characteristics as moderating variables for WMT response: For example, longer training durations in hours were associated with smaller effect sizes across studies (Teixeira-Santos et al., 2019). Note however, whereas prognostic factors are, per definition, measured and investigated on an individual-person level, the moderator analysis approach within the standard meta-analytical approach investigates modifying factors on an aggregated, study-wide level, i.e. across many individuals (e.g. mean age of participants, mean years of education). Therefore, interindividual variance of those parameters and corresponding differential training outcomes within the single-study populations are neglected in the meta-analysis of Teixeira-Santos et al. (2019). A focus on research using prognostic approaches on a single-study level would therefore substantially expand upon already existing data.

Prognostic research on treatment responsiveness after WMT has received increasing interest on a single-study level as well. However, data are inconclusive yet, as findings are highly heterogeneous and inconsistent, and prognostic approaches are often considered as an add-on analysis beyond standard effectiveness evaluations only. It seems that especially if an intervention did not yield an overall positive effect, researchers tend to exploratively analyse prognostic factors of training responsiveness. One could argue that conducting prognostic analyses on null effects might be dealing with pure noise. However, prognostic research is obliged to detangle predictors of systematic retest effects, such as practice effects or regression to the mean, from predictors of treatment response (Hingorani et al., 2013). Therefore, it is tremendously important to compare prognostic factors between a control group and the group receiving the treatment of interest (Hingorani et al., 2013). To anticipate one weakness of prognostic research in the context of cognitive interventions including WMT so far, prognostic effects are often investigated with data of the experimental group only.

Two of the most frequently investigated prognostic factors for WMT responsiveness are baseline performance in working memory or the respective cognitive outcome and general cognitive ability (e.g. Borella, Carbone, Pastore, De Beni, & Carretti, 2017; Matysiak, Kroemeke, & Brzezicka, 2019; Zinke et al., 2014). For both, inconsistent findings exist, which can be discussed within the compensation versus magnification framework (Lövdén, Brehmer, Li, & Lindenberger, 2012). Following the compensation account, individuals with lower baseline performance would show higher training benefits, because they have more room for improvement. On the contrary, the magnification hypothesis constitutes that individuals with higher abilities would benefit most, as they have more resources "to acquire, implement, and sharpen effortful cognitive strategies" (Lövdén et al., 2012). Similar inconsistent evidence exists for example for age (e.g. Borella, Carbone, et al., 2017; Borella et al., 2014; Borella, Carretti, Zanoni, Zavagnin, & De Beni, 2013; Simon et al., 2018; Zinke et al., 2014) and other demographic factors such as education (Borella, Carbone, et al., 2017; Clark, Xu, Unverzagt, & Hendrie, 2016; Matysiak et al., 2019; Mondini et al., 2016) and sex (Matysiak et al., 2019; Rahe et al., 2015; Roheger, Meyer, Kessler, & Kalbe, 2019). Furthermore, motivational processes (Kalbe, Bintener, et al., 2018; West, Bagwell, & Dark-Freudeman, 2008) and personality traits (Double & Birney, 2016; Studer-Luethi, Jaeggi, Buschkuehl, & Perrig, 2012) might constitute important individual characteristics predicting training responsiveness as well. Finally, genetic variation (Bäckman & Nyberg, 2013; Bellander et al., 2011; Brehmer et al., 2009) and brain

imaging parameters (Heinzel, Lorenz, et al., 2014; Stern, 2009) might reflect meaningful proxies for the potential to engage in cognitive plasticity following cognitive training interventions. Summarizing, a broad spectrum of potential prognostic factors to predict individual training responsiveness is discussed, however, data are inconclusive yet. Therefore, systematic reviews and meta-analyses to summarize existing evidence about prognostic factors and models of individual treatment response in the context of cognitive interventions in general and WMT in particular are urgently needed, but missing so far.

On the basis of the aforementioned considerations, the present systematic review aimed to systematically investigate prognostic factors and models for WMT responsiveness in healthy older adults. We further aimed to meta-analyse groups of "similar" prognostic effect measures to quantitatively investigate the predictive performance of the different prognostic factors. However, to anticipate one limitation of this work, data on prognostic factors after WMT was too heterogeneous and too poorly reported to conduct this meta-analysis after all.

Our systematic review question was defined using the PICOTS system as proposed by the Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS) (Debray et al., 2017; Moons et al., 2014; Riley et al., 2019). Our target population (P) consisted of healthy (i.e. absence of any neurological or psychiatric disease) older (aged ≥ 55 years) individuals. The target intervention (I) was single-domain WMT. No comparator factor (C) is being considered. Outcome variables (O) for this review are training and near-transfer effects to the domains of verbal and visuospatial working memory, as well as far-transfer effects to other cognitive domains and behavioural variables, if applicable, operationalized with objective and standardized instruments, after targeted WMT. The timing (T) of recording the relevant variables is the baseline assessment for prognostic factors and all time points of measurement for outcome variables, including follow-ups. The setting (S) was supposed to be a non-clinical one to gain prognostic information on possibilities of enhancing cognitive functioning and the prevention of cognitive decline in cognitively healthy individuals.

2 Methods

The preregistered review protocol of the present systematic review can be as accessed through https://www.crd.york.ac.uk/PROSPERO/ (ID: CRD42019142750). The reporting follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guideline for systematic reviews and meta-analyses (Moher, Liberati, Tetzlaff, & Altman, 2009). The PRISMA checklists for abstracts and systematic reviews are displayed in Supplementary Material 1.

2.1 Search Strategy

As prognostic studies are often not indexed, a broad and rather unspecific search filter was used (Riley et al., 2019). We conducted a systematic search throughout four online data bases up to October 2019: MEDLINE Ovid, Web of Science Core Collection, CENTRAL and PsycINFO. A series of keywords which were expected to appear in the title or abstract of any study containing analyses on prognostic factors or models for WMT success was created. The used keywords can be grouped into three main categories. The first category aimed to identify studies including healthy older adults as participants (e.g. "healthy elderly", "healthy aging", "older adults). The second category was used to detect a broad spectrum of interventional studies not only covering "working memory training", but also a broader spectrum of cognitive interventions (e.g. "cognitive training", "reasoning training") and even interventional studies per se (e.g. "training", "intervention"). This broad intervention category was built to ensure the search strategy to cover all kinds of WMT that are

differentially labelled in literature. The third category was included to ensure (working) memory to be a central construct of the included studies ("memory"). In addition to the systematic database search, reference lists of all relevant full texts, review articles and current treatment guidelines were hand-searched for further suitable articles. Further information and full search strings for each database can be obtained from Supplementary Material 2.

2.2 Study selection and data extraction

Title and abstract screening with predefined eligibility criteria was conducted by two reviewers (AKF and MR, or AO and MR) in Covidence Systematic Review Software (Veritas Health Innovation, available at www.covidence.org). Following, the full text articles were screened for final inclusion in the systematic review by two reviewers (AO and MR). If a full text was not available online, we contacted the corresponding authors and asked to provide the full text publication within 2 weeks of time. If no consensus was reached between the two reviewers (AO and MR), the plan was to discuss the case with a third author (NS) until a final consensus was reached, however, this option was not needed. Relevant data considering general study characteristics (e.g. participants' demographics, WMT features) and prognostic factor and/or model analyses was independently extracted by two reviewers (AO and MR) according to the CHARMS checklist (Moons et al., 2014).

2.3 Eligibility Criteria

Inclusion criteria for our systematic review were (i) full text research article publication until October 2019 in a peer-reviewed journal in English or German, (ii) inclusion of healthy older individuals aged ≥ 55 years without any neurological and/or psychiatric diseases including cognitive impairment (mild cognitive impairment or dementia), as well as uncorrected seeing or hearing impairments assessed via self-report, (iii) investigation of prognostic factors and/or models for training responsiveness in terms of direct training and near-transfer effects to verbal and visuospatial working memory, as well as far transfer effects to other cognitive domains and behavioural variables, operationalized with objective and standardized instruments, after targeted WMT.

Age of ≥ 55 years was chosen as a cut-off, as we on the one hand wanted to provide an objective age cut-off for individuals within the included studies, and on the other hand did not want to exclude studies including healthy older individuals just below the frequently used cut-off of ≥ 60 years (e.g. Sala et al., 2019; Soveri et al., 2017). Targeted WMT was defined as a cognitive training either computerized, with paper-pencil tasks, or mixed, which is administered either on personal devices or in individual- or group settings with a minimum of 2 training sessions. When multi-domain trainings were examined, working memory had to be the main component of the program (defined as being the main target in at least 80% of the exercises). Verbal and visuospatial working memory, i.e. direct training and near-transfer effects were defined as primary outcomes, with direct training effects constituting effects in trained working memory tasks over the course of training and near-transfer effects constituting effects in untrained working memory tasks. Other cognitive far-transfer outcomes (i.e. effects in untrained cognitive domains, e.g. global cognition, memory, fluid intelligence, executive functions, attention) and clinical and patient-centered outcomes (e.g. depressive symptoms, quality of life) were considered as secondary outcomes. Both primary and secondary outcomes needed to be assessed with established and objective psychometric instruments.

For the systematic review, we considered all prognostic factors (e.g. sociodemographic factors, cognitive abilities at the entry of the training, brain imaging parameters, genetic parameters, personality traits, training-related characteristics), which investigate critical aspects of WMT responsiveness. As outlined in the introduction, a prognostic factor is defined as any measure that, among people with a given condition (e.g. the process of aging), is associated with a subsequent

outcome (e.g. changes in cognition after certain interventions) (Riley et al., 2013). Prognostic factor finding studies aim at establishing one or several variables as independent prognostic factors associated with an outcome. In contrast, prognostic model studies identify more than one prognostic factor, assign relative weights to each prognostic factor, and estimate the model's predictive performance through calibration and discrimination (Moons et al., 2009). We included all studies investigating prognostic factors and/or prognostic models regardless of whether or not significant general training effects and/or significant relationships between prognostic factors and training responsiveness were found.

2.4 Quality Assessment

Using the Quality in Prognosis Studies (QUIPS) checklist (Hayden, van der Windt, Cartwright, Côté, & Bombardier, 2013), risk of bias of the included studies was examined independently by two reviewers (AO and MR) across six domains: study participation, study attrition, prognostic factor measurement, outcome measurement, adjustment for other prognostic factors, statistical analyses and reporting. Each domain was overall rated with high, moderate, or low risk, depending on the rating in the corresponding items. A detailed description of the QUIPS checklist including each item and the overall judgment rules implemented by the two reviewers is presented in Supplementary Material 3. Instead of using two different risk of bias assessment tools (QUIPS, Hayden et al., 2013, for prognostic factor finding studies; PROBAST, Prediction model Risk of Bias Assessment Tool, Wolff et al., 2019, for prognostic model studies), risk of bias of both prognostic factor finding and prognostic model studies was assessed with the QUIPS tool to get a comparable risk of bias rating.

2.5 Data Analysis

Initially, and as stated in the pre-registration of the study, we aimed to meta-analyse groups of "similar" prognostic effect measures with a random effects approach to investigate the predictive performance of the different prognostic factors. However, after data extraction, we had to ascertain that data on prognostic factors after WMT was too heterogeneous and too poorly reported to conduct this meta-analysis. The main reason was that we were not able to compute comparable effect size measures (e.g. odds ratios, hazard ratios) to meta-analyse the prognostic effects reported in the studies due to the fact that either data was not reported and could not be assessed within studies, or data was not consistent enough across studies to pool the results. Therefore, the systematic review focussed on the qualitative directionality of the prognostic effects reported in the included studies rather than their magnitude.

3 Results

3.1 Study Flow

12,966 records were identified through our database search. After removing duplicates, titles and abstracts of 9,583 records were screened for eligibility. As prognostic analyses are often not indexed, title and abstract screening focused on the content-related criteria "healthy older individuals" and "working memory training". Thus, 138 full texts were screened for eligibility. Finally, $n = 16$ studies were included in the present systematic review. For details on study flow and reasons for exclusion, see Figure 1 (PRISMA Flow Diagram).

3.2 Descriptive Characteristics of Included Studies

Within the 16 studies, n = 675 healthy older individuals with a mean age ranging from 63.0 years (Brehmer et al., 2011) to 86.8 years (Zinke, Zeintl, Eschen, Herzog, & Kliegel, 2012) were investigated, of which 63% were women. Years of formal education ranged from a mean of 5.72 years (Borella et al., 2013) to 18.65 years (Tusch et al., 2016). Throughout the studies, different training regimes that varied in terms of setting, number of sessions, total time of training, and training tasks were applied. The number of training sessions ranged from three (Borella, Carbone, et al., 2017; Borella et al., 2014; Borella, Carretti, Meneghetti, et al., 2017; Borella, Carretti, Sciore, et al., 2017; Borella et al., 2013; Brum, Borella, Carretti, & Sanches Yassuda, 2018) to 25 (Brehmer et al., 2011; Matysiak et al., 2019; McAvinue et al., 2013; Simon et al., 2018; Tusch et al., 2016) with a total time of training ranging from 105 minutes (Brum et al., 2018) to 1000 minutes (Tusch et al., 2016). 44% of trainings addressed verbal working memory only and 50% followed a mixed approach addressing both verbal and visuospatial working memory. Only one study conducted a multi-domain WMT, as next to working memory tasks one executive control task was included within the training regime (Zinke et al., 2014). All training regimes were conceptualized as adaptive, except for those studies, in which adaptivity was investigated as a prognostic factor for WMT responsiveness (Brehmer et al., 2011; Simon et al., 2018; Tusch et al., 2016; Weicker et al., 2018).

In total, nine studies applied digital WMT: Four studies used commercially available, digital WMT programs (Cogmed and WOME/ RehaCom®) (Brehmer et al., 2011; Simon et al., 2018; Tusch et al., 2016; Weicker et al., 2018), three studies used a digital n-back training (Heinzel, Lorenz, et al., 2014; Heinzel, Riemer, et al., 2014; Matysiak et al., 2019), and two used a study-individual composition of digital WMT tasks (Borella et al., 2014; McAvinue et al., 2013). Five studies used a WMT with the Categorization Working Memory Span (CWMS) Task based on audio-recordings (Borella, Carbone, et al., 2017; Borella, Carretti, Meneghetti, et al., 2017; Borella, Carretti, Sciore, et al., 2017; Borella et al., 2013; Brum et al., 2018), however, all of these studies were conducted by the same group of researchers. Only two studies used paper-pencil WMT (Zinke et al., 2012; Zinke et al., 2014). For details on study, participants, and training characteristics, see Table 1.

3.3 Reporting quality and risk of bias

Table 2 reports the risk of bias per study across six domains evaluated with the QUIPS checklist (Hayden et al., 2013). A detailed risk of bias assessment on a single item level rather than QUIPS domain ratings can be obtained from the corresponding author. Important information is lacking throughout many of investigated studies, especially regarding the domains study attrition, study confounding and statistical analysis and reporting. Most notably, the appropriate selection of the analysis plan and reporting of both the statistical analyses and results is often fragmentary. Only for the domains of prognostic factor measurement and outcome measurement the majority of studies was rated with low risk. Summarizing, the reporting quality was partly insufficient and results should be interpreted cautiously.

Unfortunately, the initially planned meta-analysis could not be performed as the applied analytical approaches, as described below, were too heterogeneous and the reported results did not allow to compute comparable effect size measures (e.g. odds ratios, hazard ratios) across studies to meta-analyse the prognostic effects. Therefore, only a systematic review focusing on the directionality of prognostic effects rather than their magnitude was performed.

3.4 Prediction analyses and outcome measures

Seven of the 16 prognostic studies used more than one prediction analysis account to predict WMT responsiveness (one study included both a prediction model and a factor finding approach; six

studies included more than one factor finding approach, i.e. investigated the prognostic value of one or several variables with at least two different approaches). Five studies investigated prediction models, three of which used hierarchical regression analyses (Borella, Carretti, Sciore, et al., 2017; Heinzel, Lorenz, et al., 2014; Zinke et al., 2014) with change scores or relative change scores as dependent variables. One study used a Bayesian modelling approach (Borella, Carbone, et al., 2017) and one Linear Mixed Effect Modelling (Simon et al., 2018), both with time as one predictor, therefore abandoning the use of change scores as dependent variable. Ten studies were factor finding studies, including a total of 18 factor finding analysis approaches: seven used a generalized linear model approach (e.g. ANOVA) (Borella et al., 2014; Brehmer et al., 2011; Brum et al., 2018; Heinzel, Riemer, et al., 2014; Tusch et al., 2016; Weicker et al., 2018; Zinke et al., 2012), one used ANCOVA (Borella, Carretti, Meneghetti, et al., 2017), five Pearson correlations (Brehmer et al., 2011; Heinzel, Lorenz, et al., 2014; McAvinue et al., 2013; Tusch et al., 2016; Zinke et al., 2012), one linear regressions (Weicker et al., 2018) and one Linear Mixed Models (Matysiak et al., 2019). Three studies used a (descriptive) comparison of effect sizes (Borella et al., 2014; Borella et al., 2013; Brum et al., 2018). For the generalized mixed model approach, 71% used time as a predictor and only 29% used raw or standardized change scores as dependent variable. For the ANCOVA, the post-test score was used as dependent variable. Pearson correlations and linear regressions used (standardized) change scores as dependent variables, for the Linear Mixed Model, time was used as predictor. None of the studies compared prognostic factors or models between the trained group and a passive control group, i.e. they analysed data of trained groups only. Summarizing, even though prediction approaches were highly heterogeneous, analyses were comparable within the different approaches.

We defined verbal and visuospatial working memory, i.e. direct training and near-transfer effects, as primary outcomes. Most of the included studies distinguished between these two domains, however, four studies did not (Brehmer et al., 2011; Simon et al., 2018; Weicker et al., 2018; Zinke et al., 2012), and four studies addressed verbal working memory only (Heinzel, Lorenz, et al., 2014; Heinzel, Riemer, et al., 2014; Matysiak et al., 2019; Tusch et al., 2016). Three of the 16 included studies (18.8%) investigated direct training effects (i.e. effects in trained tasks) only (Heinzel, Lorenz, et al., 2014; Heinzel, Riemer, et al., 2014; Matysiak et al., 2019). The majority of studies (62.5%) investigated a combination of direct training, near-transfer (i.e. untrained working memory tasks), and several far-transfer measures, defined as secondary outcomes in our systematic review. Frequently investigated far-transfer cognitive domains were executive functions (including verbal fluency, reasoning, inhibition, set-shifting, and executive control), processing speed, (short-term) memory, and fluid intelligence. Only one study investigated non-cognitive outcomes (anxiety and depression) (McAvinue et al., 2013). Only three of the included studies (18.8%) did not apply a prognostic approach for at least one direct training outcome and instead focused on near- and far-transfer effects only (McAvinue et al., 2013; Simon et al., 2018; Tusch et al., 2016). Most studies used objective and standardized neuropsychological assessment tools. Others, for example studies assessing (verbal) working memory by n-back tasks (25%), compared n-back task levels within different points of time or used indexes from signal detection theory (Heinzel, Lorenz, et al., 2014; Heinzel, Riemer, et al., 2014; Matysiak et al., 2019; Tusch et al., 2016). For details on prediction analyses and outcomes, see Table 3 and Supplementary Material 4.

3.5 Predictor variables and prediction results

Several different predictors for WMT responsiveness were investigated, including individual-related sociodemographic factors (e.g. age, sex, education), cognitive variables (baseline performance, intelligence, processing speed), and biological factors (genes, brain metabolism), as well as training-

related factors (e.g. adaptivity, dose of training). 13 analysis approaches investigated individual-related prognostic factors only, two analysis approaches investigated a combination of individual- and training-related characteristics, and eight analysis approaches investigated training-related characteristics only as predictors for WMT responsiveness. Results of the prognostic analyses are reported in Table 3. As in most cases the direction of predictor effects did not vary systematically between single outcome variables, and within prognostic factor finding versus prognostic model studies, we decided to not further distinguish between different outcome variables and prognostic factor finding versus prognostic model studies, but indicate if prognostic effects were found for direct training and/or near- and/or far-transfer effects only. Described patterns of prognostic effects only reflect frequencies of observed prognostic relationships and do not take into account risk of bias and further methodological shortcomings of the underlying studies.

Age was investigated in four of five prognostic model studies and three of 18 factor finding approaches. With only few exceptions for single outcome measures reporting positive or non-significant relationships, age was consistently found to be a negative predictor for WMT responsiveness across direct training as well as both near- and far-transfer effects, i.e. younger participants benefitted more from the training than older participants independent of outcome measures. Note, however, that age as a continuous variable was dichotomized into young-olds vs. old-olds for 3 analytical approaches investigating age as a predictor for WMT responsiveness (Borella et al., 2014; Borella et al., 2013; Simon et al., 2018).

Education was investigated within two prognostic model and two factor finding approaches. Education most frequently constituted a negative predictor for direct training as well as near- and far-transfer effects (Borella, Carbone, et al., 2017; Heinzel, Lorenz, et al., 2014), however, some analyses do not yield a significant relationship at all (Matysiak et al., 2019; Tusch et al., 2016). Whereas education was treated as a continuous variable in most studies, Matysiak et al. (2019) dichotomized the variable for their analysis. Sex was investigated in one prognostic model and one factor finding approach, but was not found to be a significant predictor of WMT responsiveness in direct training effects (Heinzel, Lorenz, et al., 2014; Matysiak et al., 2019) and was not investigated in any prognostic approach on near- and/or far-transfer measures).

Baseline performance in working memory tasks and/or outcome measures was the most frequently investigated prognostic factor (four of five prognostic model studies and five of 18 factor finding approaches). For both near- and far-transfer outcomes, baseline working memory and/or baseline performance in outcome measure was consistently found to be a negative predictor for WMT responsiveness (Borella, Carbone, et al., 2017; Borella, Carretti, Sciore, et al., 2017; Zinke et al., 2012), i.e. individuals with lower performance at baseline improved more from WMT than individuals with higher baseline performance. However, for analyses on direct training effects, heterogeneous results appear with some analyses indicating baseline working memory and/or baseline performance in outcome measure to be a positive predictor (Brehmer et al., 2011; Heinzel, Lorenz, et al., 2014; Matysiak et al., 2019; Weicker et al., 2018), i.e. individuals with higher baseline performance in training tasks achieving higher WMT task gains than individuals with lower baseline performance. Baseline performance as a continuous variable, was dichotomized into high- vs. low-performers by median split in two of the analytical approaches (Matysiak et al., 2019; Zinke et al., 2012).

Intelligence was investigated within two of five prognostic model studies and one of 18 factor finding approaches. For direct transfer effects, the prognostic value remains unclear (Borella, Carbone, et al., 2017; Zinke et al., 2014). Furthermore, whereas there does not seem to be a significant predictive value when intelligence is investigated as a prognostic factor for near-transfer effects (Tusch et al., 2016; Zinke et al., 2014) or evidence points to different prognostic directions

(Borella, Carbone, et al., 2017), a different pattern emerges for far-transfer effects: If significant, for the majority of far-transfer effect outcomes, the analyses indicate intelligence to be a positive predictor of gains after WMT (Borella, Carbone, et al., 2017; Zinke et al., 2014), i.e. individuals with higher intelligence show larger far-transfer effects after targeted WMT than individuals with lower intelligence.

In the only study (prognostic model and prognostic factor) investigating a functional imaging parameter as predictor for WMT gains, individuals with a BOLD response pattern more similar to younger adults (i.e. higher load-dependent network Delta scores) showed higher direct WMT gains (Heinzel, Lorenz, et al., 2014). Only one study investigated a genetic factor, yielding carriers of the Val/Val catechol-O-methyltransferase (COMT) genotype to show less direct training effects after WMT than carriers of any Met COMT genotype (Heinzel, Riemer, et al., 2014).

With regard to training-related prognostic factors, prognostic effects of dose of training (investigated within two studies) were mixed for both near- and far-transfer effects (Brum et al., 2018; McAvinue et al., 2013), only marginally comparable between studies because of different prognostic factor operationalizations, and not investigated for direct training effects. Adaptivity was investigated within four studies and, if significant, showed to be a positive predictor for WMT responsiveness (Brehmer et al., 2011; Simon et al., 2018; Weicker et al., 2018), with adaptive training regimes yielding better results than non-adaptive training regimes, especially for near-transfer effects.

4 Discussion

This systematic review is the first one evaluating prognostic factors and models for WMT responsiveness in healthy older adults. Within the 16 studies meeting our inclusion criteria, five prognostic model approaches and 18 factor finding approaches were included. One of the main findings is that methodological and reporting quality of prognostic research within the evaluation of WMT regimes in healthy older adults is often insufficient, therefore, no meta-analysis could be conducted and no clear conclusions can be drawn from the systematic review. Age, education, intelligence, and baseline performance in working memory or other cognitive domains were frequently investigated predictors across studies. However, given the methodological shortcomings of the included studies, emerging patterns of prognostic effects across direct training as well as near- and far-transfer effects will have to survive sound methodological replication in future attempts to promote precision medicine approaches in the context of WMT.

First, our findings will be discussed within the methodological framework of prognostic research, secondly, they will be related to the theoretical framework of cognitive aging and embedded into other prognostic research literature in the field of cognitive interventions, and thirdly, they will be linked to findings from a prognostic review on memory trainings in healthy older adults (Roheger, Folkerts, Krohm, Skoetz, & Kalbe, 2020).

4.1 Methodological considerations

Several methodological considerations and implications can be derived from the present systematic review. First of all, it has confirmed that prognostic research in the area of WMT in healthy older adults is not yet fully established and rather premature. The prognostic framework is usually not indexed and the specific mention of the prognostic approach in titles or abstracts is limited as well (Riley et al., 2019). For example, within our included studies, only 5 studies used a prediction-related terminology in their titles (Borella, Carbone, et al., 2017; Heinzel, Lorenz, et al., 2014; Heinzel, Riemer, et al., 2014; Matysiak et al., 2019; Zinke et al., 2014).

Furthermore, large heterogeneity appears throughout the included studies with regard to study design (e.g. randomized controlled trials vs. cohort studies vs. post-hoc analyses) and the applied analytical approaches. The applied analytical approaches did not only differ widely per se, but have differing suitability to answer the question of "who benefits most" from WMT regimes in healthy older adults. In general, a prognostic factor is defined as any measure that, among people with a given condition, is associated with a subsequent outcome (Riley et al., 2013), therefore implying at least some kind of a causal relationship. The majority of studies in our systematic review, however, used group-comparisons (e.g. by ANOVA, *t*-test, comparison of effect sizes) to investigate the influence of a group-characteristic on a given outcome. Despite the fact that these approaches can only state whether the compared groups differ from one another and not whether the investigated group characteristic linearly correlate with or even causally predict the investigated outcome, another important point needs to be highlighted: Whereas some investigated prognostic factors are innately categorical (e.g. sex, training modality, adaptivity), originally continuous predictors (e.g. age, baseline performance) were frequently dichotomized into artificial groups, for example young-olds vs. old-olds (Borella et al., 2014; Borella et al., 2013; Simon et al., 2018), and high- vs. low-performers (Matysiak et al., 2019; Zinke et al., 2012). Dichotomization of both dependent and independent variables is strongly discouraged as it results in a loss of information, possible misunderstandings of actual continuous relationships, and a severe loss of power (Dawson & Weiss, 2012; Fernandes, Malaquias, Figueiredo, da Rocha, & Lins, 2019; Moreau, Kirk, & Waldie, 2016).

Another frequently used analytical approach was the computation of correlation coefficients between predictor variables and change scores in outcome measures after WMT. However, no causal interferences can be derived from correlation analyses (Bewick, Cheek, & Ball, 2003). Furthermore, correlations for example between baseline performance and change scores (which is obtained by subtracting baseline performance from post-training performance), are less more than pure statistical artefacts (Smoleń, Jastrzebski, Estrada, & Chuderski, 2018). Smoleń et al. (2018) discuss that unfortunately, even more advanced methods such as multiple regressions and linear mixed models do not guarantee the correct assessment of relationships between predictor variables and respective outcomes. According to the authors, the only correct method would be to use direct modelling of correlations between latent true measures and gain by structural equation modelling (Smoleń et al., 2018). Future research on prognostic factors regarding (working memory) training responsiveness should apply advanced statistical methods such as latent difference score models or growth curve analyses as highly flexible statistical approaches from the structural equation modeling background. On the one hand, this would allow to circumvent several statistical fallacies clinical trial data often include such as violations of multivariate normality assumptions, non-linear change trajectories, and missing data patterns (Newsom, 2015). On the other hand, it would allow to explore the (statistical) properties of change through training without actually calculating change scores and with highly flexible options to model interdependencies between several variables (Smoleń et al., 2018).

In this context one immense problem arises within prognostic research on cognitive intervention programs per se and WMT in particular: a lack of statistical power due to small sample sizes. Prognostic research requires large sample sizes with a representative distribution of individuals' characteristics and values across the prognostic factors of interest. Especially for (cognitive) training studies, researchers are confronted with the challenge to overcome a self-selection bias to not only engage highly educated, active and motivated individuals within their trials (Oswald, Gunzelmann, Rupprecht, & Hagen, 2006; Schubert, Strobach, & Karbach, 2014). As prognostic research in this field often arises as an (explorative) add-on or post-hoc analysis of former data from randomized controlled trials, sample size calculations at the stage of study design (if present at all) do only take into account the sample size needed to evaluate effectiveness of a training regime (by comparing the experimental group against at least one control group). For future research

in the field of personalized prevention and treatment approaches for healthy aging, we encourage to emphasize the outstanding importance of prognostic research by focusing on the prognostic aim already during study design.

Importantly, as already discussed in the introduction, prognostic analyses should always include data of at least one control group as well to detangle predictors of specific treatment response from general prognostic factors of retest effects such as practice effects and regression to the mean (Hingorani et al., 2013). None of the studies included in this systematic review followed this recommendation. Therefore, the identified prognostic relationships might represent systematic relationships, however, they might exist in both treated and untreated individuals, and, therefore, not represent true predictors of treatment response.

Beyond that, however, the large body of data on WMT effectiveness for healthy older adults bears the enormous potential of post-hoc prognostic analyses, for example as executed by Borella, Carbone, et al. (2017). Within the tradition of evaluating similar WMT regimes, over the years, several randomized controlled trials to investigate efficacy of similar training regimes were carried out in this study group. As Borella, Carbone, et al. (2017) recognized large variability in the effectiveness of WMT across individuals on the one hand, and large heterogeneity across results on earlier investigations on the influence of individual characteristics on training outcomes on the other hand, they merged the data of four earlier training studies (Borella, Carretti, Riboldi, & De Beni, 2010; Borella, Carretti, Sciore, et al., 2017; Borella et al., 2013; Carretti, Borella, Zavagnin, & de Beni, 2013) to investigate individual's characteristics related to WMT gains in a larger sample. In other words, they conducted a tiny-scale individual participant data (IPD) meta-analysis, the gold-standard for meta-analytical approaches. At this point, it should be noted, that Borella, Carbone, et al. (2017) included data of participants from the training groups of Borella et al. (2013) and Borella, Carretti, Sciore, et al. (2017), two studies included in our systematic review as well. Therefore, prognostic results of these three studies are not fully independent. However, we did not exclude the two earlier works, as the exclusion would not have changed the results on the (qualitative) directional prognostic effects. For a future IPD meta-analysis, IPD data of either the four mentioned studies or Borella, Carbone, et al. (2017) should be included only.

Regarding the used analytical approaches and results of this review, it should further be mentioned that the recommendation, to focus on adjusted results to reveal whether a certain index factor contributes independently and above other prognostic factors (Riley et al., 2019), could not be met entirely: most of the included studies in this review investigated only one prognostic factor per analysis. However, as established prognostic factors did not (yet) exist in the context of WMT responsiveness, analytical approaches excluding possibly important confounding variables are (at least in parts) comprehensible as well. For future prognostic research in this field, however, we recommend to include baseline performance and age as a minimum set of control variables when investigating further prognostic factors.

4.2 Prognostic factors for working memory training responsiveness

Several different predictors for WMT responsiveness were investigated, including individual-related sociodemographic factors (e.g. age, sex, education), cognitive variables (baseline performance, intelligence), and biological factors (brain metabolism, genes), as well as training-related factors (e.g. adaptivity, dose of training). Given the methodological shortcomings of the included studies discussed above, no clear conclusions regarding prognostic effects can be drawn. Emerging patterns based on frequently observed prognostic effects will have to survive sound methodological replication in future attempts to promote precision medicine approaches in the context of WMT.

Some inconsistent findings might be due to statistical and psychometric artefacts, uncontrolled extraneous influences, or the absence of convincing robust prognostic relationships at all. Nevertheless, we would like to provide a contextual framework for the discussion of possible predictors for WMT responsiveness beyond pure methodological issues.

The most frequently investigated predictor was baseline performance. Despite the many different statistical approaches and poor reporting quality in most studies, baseline performance was, with exceptions for direct-training effects only (Brehmer et al., 2011; Heinzel, Lorenz, et al., 2014; Matysiak et al., 2019; Weicker et al., 2018), identified as a negative predictor, i.e. individuals with lower baseline performance are the ones that benefit most from WMT in terms of performance on neuropsychological tests in the domains of working memory and other cognitive functions (e.g. executive functions, short-term memory). Therefore, most inconsistencies regarding the directionality of the prognostic effect of baseline performance could be elucidated when taking a look at the operationalization of the dependent variables. The finding of baseline performance being a negative predictor for cognitive intervention responsiveness is also common for targeted memory trainings (Roheger et al., 2020), as well as other cognitive intervention approaches such as multidomain cognitive trainings (López-Higes et al., 2018; Roheger et al., 2019; Whitlock, McLaughlin, & Allaire, 2012). However, opposing findings exist as well, indicating that higher baseline performance might be indicative for cognitive intervention success (Fairchild, Friedman, Rosen, & Yesavage, 2013; Willis & Caskie, 2013). However, given the lack of comparisons of prognostic factors between WMT and control groups within the included studies, the frequently observed negative associations between baseline performance and change through training might simply represent effects of regression to the mean (Smoleń et al., 2018). This statistical artefact causes negative correlations between baseline performance and gain by noisy repeated measurements, where extreme values at the first point of time tend to be closer to the mean at the second point of time without reflecting real change (Smoleń et al., 2018).

Nevertheless, baseline performance as a predictor for training responsiveness can be discussed within the compensation versus magnification framework (Lövdén, Bäckman, Lindenberger, Schaefer, & Schmiedek, 2010; Lövdén et al., 2012). Following this account, individuals with lower baseline performance would show higher training benefits, because they have more room for improvement, whereas individuals with higher baseline performance already perform at ceiling, leaving less room for improvement. Improvements across individuals performing less optimal at baseline might therefore represent some kind of flexibility rather than plasticity. According to Lövdén et al. (2010), flexibility represents “the capacity to optimize the brain’s performance within current structural constraints, using the available range of existing representational states”. Beyond this flexibility, plasticity denotes the capacity for extending the range of representational states, where flexibility then operates. This understanding of plasticity, however, fits better with the magnification hypothesis, constituting that individuals with higher cognitive abilities would benefit most, as they have more resources “to acquire, implement, and sharpen effortful cognitive strategies” (Lövdén et al., 2012).

Within our systematic review, we also found hints for this dualism between compensation versus magnification or rather flexibility versus plasticity. Whereas our findings regarding baseline performance in neuropsychological test measures might rather reflect mechanisms following the compensation account, our findings regarding age as a possibly negative predictor and intelligence as a possibly positive predictor for WMT responsiveness are more interpretable in terms of the magnification account. Higher (crystallized) intelligence might constitute the required “hardware” to utilize the possibilities given by WMT to extend the cognitive repertoire, and, in the broadest sense, reflecting cognitive plasticity. This perspective is strengthened considering our finding that

intelligence seems to be a positive predictor for gains after WMT for far-transfer effects only. Whereas lower baseline performance might be predictive for both near- and far-transfer effects (interpreted in terms of the compensation account and flexibility: if there is room for improvement, performance will be optimized by training), higher cognitive abilities might be especially beneficial for far-transfer effects, i.e. to transfer direct training effects to untrained cognitive domains. The magnification account might additionally be able to explain our finding that baseline performance in trained tasks sometimes emerged as a positive predictor for direct training effects. As most WMT regimes adapted their difficulty to user performance across the course of training and no ceiling effects could be expected, higher initial levels might represent general cognitive ability rather than task specific baseline, and participants with higher initial levels in training tasks might be more able to utilize the whole potential of the training regime.

The second most frequently investigated predictor was age, indicating that older individuals might benefit less from WMT than younger individuals, even within the cohort of healthy older adults above the age of 55. Age might be a proxy for the course of the interplay between neural and cognitive plasticity, which yields a higher potential for plastic changes in younger age than in old-old age (Burke & Barnes, 2006; Greenwood & Parasuraman, 2010; Li, 2013). Due to age-related reductions in processing resources (Paraskevoudi et al., 2018; Park & Bischof, 2013), the ability to engage in plastic changes after WMT might be reduced in older age. This was already reflected in an early meta-analysis on moderators of memory training effects (Verhaeghen, Marcoen, & Goossens, 1992). However, findings in contemporary cognitive intervention literature diverge and either report no significant relationship (Roheger et al., 2019; Willis & Caskie, 2013), positive relationships (i.e. the older the individual, the more benefits) (Brooks, Friedman, Pearman, Gray, & Yesavage, 1999), or negative relationships (i.e. the younger the individual, the more benefits) (Fairchild et al., 2013). In terms of differential prognostic effects for different training regimes (e.g. WMT versus memory training), this will be further discussed below.

The only study investigating brain imaging parameters as predictors for WMT responsiveness strengthen the finding of our systematic review that age might be a negative predictor for positive training responsiveness: Heinzel, Lorenz, et al. (2014) found a more "youth-like" BOLD response pattern in healthy older adults to be predictive of increased working memory performance after training. This youth-like response pattern is reflected in a higher load-dependent working memory network Delta score, indicating that both high working memory network efficiency (represented by decreased activation during low-level tasks) and high working memory network capacity (represented by increased activation during high-level tasks) are related to plasticity (Barulli & Stern, 2013). This BOLD response pattern has also been discussed as a biomarker for cognitive reserve (Stern, 2009). Against this backdrop, one could hypothesize that cognitive reserve and brain reserve constitute higher-order predictors for WMT success and are operationalized by several different proxies within the existing prognostic research approaches (Stern et al., 2018).

Within the cognitive reserve framework, it is not uncommon to find education alone as a proxy for this construct (Stern, 2002; Stern et al., 2018; Valenzuela & Sachdev, 2006). In our systematic review, we found a tendency of education being a negative predictor of WMT responsiveness. In cognitive intervention research, it is discussed that cognitive interventions might be able to diminish the cognitive reserve disadvantage of less-educated older adults (Clark et al., 2016; Mondini et al., 2016), thereby leading to more training-related gains. As this might appear counterintuitive at first, it is important here to differentiate between brain reserve and lifetime proxies of cognitive reserve such as education, occupational attainment, and leisure time activities (Stern, Barnes, Grady, Jones, & Raz, 2019). A higher cognitive reserve is commonly associated with less

cognitive deficits given the same brain pathology (Hoenig et al., 2019; Wilson et al., 2013). Following, two individuals with similar cognitive functioning but different educational backgrounds might also differ in their brain pathology, i.e. the individual with higher education might already show a higher level of brain pathology compared to the individual with lower education, which in turn comes down to lower levels of brain reserve for individuals with higher education. Therefore, for the individual with lower education, even though lifetime cognitive reserve is lower, brain reserve might be higher, which corresponds to a better hardware to adapt training benefits.

Only one study investigated a genetic factor as predictor for WMT responsiveness in healthy older adults (Heinzel, Riemer, et al., 2014), revealing carriers of the Val/Val COMT genotype, which is associated with reduced prefrontal dopamine metabolism, to benefit less from WMT than carriers of any Met COMT genotype. The COMT genotype affects prefrontal dopamine metabolism which is itself related to cognitive plasticity (higher prefrontal dopamine metabolism = more cognitive plasticity) (Diamond, 2007; Frias et al., 2005). Furthermore, previous research indicated that advantageous dopamine-related genes are critically involved in working memory performance and the ability to benefit from WMT (Bäckman & Nyberg, 2013; Bellander et al., 2011; Brehmer et al., 2009), which further strengthens the finding of Heinzel, Riemer, et al. (2014) that these relationships are also present in healthy older adults.

We did not find a consistent influence of sex on responsiveness to WMT in healthy older adults, even though some kind of "sex-specific plasticity" and following sex-specific differences between training responsiveness to different cognitive domains are proposed in literature (Beinhoff, Tumani, Brettschneider, Bittner, & Riepe, 2008; Rahe et al., 2015; Roheger et al., 2019). Note, however, that sex as a prognostic factor for WMT responsiveness was investigated in two studies with direct training effects as dependent variable only. Therefore, no final conclusions can be drawn. Even though motivational factors and personality traits are discussed to play a significant role in predicting responsiveness to general cognitive interventions (Colquitt, LePine, & Noe, 2000; Double & Birney, 2016; Kalbe et al., 2018; Studer-Luethi et al., 2012; West, Bagwell, & Dark-Freudeman, 2008), they were not yet investigated as prognostic factors within the WMT context.

Summarizing possible prerequisites for WMT responsiveness, we hypothesize that there has to be room for improvement (i.e. lower baseline performance) to engage in training-related cognitive flexibility, but also sufficient "hardware" (e.g. age, intelligence, brain metabolism, genetic variation) to engage in training-related cognitive and neural plasticity. Again, it needs to be highlighted that the body of evidence (so far) is too weak to draw clear conclusions. Even though some findings fit well into the compensation versus magnification account and the cognitive reserve framework, future studies of high methodological quality will have to replicate those findings.

Regarding dose of training as one training-related prognostic factor investigated in the context of WMT responsiveness, results were mixed and are in accordance with heterogeneous results in literature. For example, Teixeira-Santos et al. (2019) identified shorter compared to longer training durations to be beneficial for training outcome. However, they discuss this finding to be unexpected and influenced by confounding factors such as the type of outcome variable and highly heterogeneous training durations that impede comparability between studies. All of the included studies in our review implemented an adaptive training regime, where the task difficulty adapted to user performance. Four studies compared adaptive versus non-adaptive WMT regimes, with adaptivity emerging as a positive predictor for training responsiveness. Adaptivity of trained task difficulty is discussed to contribute to the maintenance of training motivation and the avoidance of underchallenging and overstraining participants during training (Weicker et al., 2016). However, some studies did not find beneficial effects of implementing individually adaptive training regimes (von Bastian & Eschen, 2016).

Only one study within our systematic review used a multi-domain training. Zinke et al. (2014) included an executive control task next to several working memory tasks within their WMT regime. Executive control might, however, strongly be dependent on working memory (Chai et al., 2018). Even though, we cannot evaluate the contribution of single training tasks or the training of single domains to the overall prognostic effects, we conclude that this exception from targeted WMT does not constitute a danger for the validity of our findings regarding WMT responsiveness.

4.3 Working memory training versus memory training

Just recently, a systematic review on prognostic factors of memory improvements after memory training using a similar systematic review technique has been published (Roheger et al., 2020). Roheger et al. (2020) identified further methodological short-comings of prognostic research in the context of memory training and, on a content-related level, more vulnerable individuals (e.g. lower baseline performance, higher age) to benefit most from memory training. They also identified several "hardware" factors (e.g. hippocampal volume, genetic variation in Apolipoprotein-E-(apoE)4) as prognostic factors. Primarily, however, the direction of age as a prognostic factor seems to differ between the two training regimes.

We hypothesize this difference to be due to the different cognitive training approaches investigated. Memory training, as investigated by Roheger et al. (2020), can be referred to as a strategy-based training, whereas WMT can be referred to as a process-based training (Lustig et al., 2009; Teixeira-Santos et al., 2019). Whereas strategy-based trainings focus on the application of specific strategies to a task were the target population typically does poorly, process-based trainings focus on tasks that load on a specific cognitive function, however, without explicit strategy training (Lustig et al., 2009). Thereby, process-based trainings are believed to produce more transfer effects to untrained domains, as untrained cognitive functions might depend on the targeted cognitive domain (Lustig et al., 2009; Teixeira-Santos et al., 2019). This difference in the conceptualization of memory training versus WMTs, however, implicates different levels of cognitive demands that have to be met in order to benefit from the trainings. Given the higher cognitive demands of WMT, we hypothesize that younger individuals might benefit more, as their hardware potential to engage in neural and cognitive plasticity is higher. Older individuals, however, might be less able to engage in neural plasticity, but might therefore rather benefit from strategy-based training approaches optimizing their cognitive performance within a given structural constraints in terms of flexibility (Lövdén et al., 2010; Lövdén et al., 2012). In the framework of Lövdén et al. (2012), WMT gains equal practice gains that are related to plasticity and better fit the magnification model, whereas memory training gains equal instruction gains that are related to flexibility and better fit the compensation account. Further research is necessary to proof this concept, but we are convinced that these findings highlight the urgent need for personalized cognitive prevention and intervention methods to counteract cognitive decline at best for every individual.

Another systematic review and meta-analysis on prognostic factors and models of cognitive and behavioural changes after multidomain cognitive training in healthy older adults is still ongoing (preliminary Prospero ID 147531). Those findings, in combination with findings of the present systematic review and Roheger et al. (2020), will further contribute to the understanding of which cognitive interventions yield best outcomes for which individual. Furthermore, the discussion around precision medicine in the context of cognitive interventions can be taken to a whole new level, if one would not only consider the cognitive domain trained (or the combination of domains), but also the nature of the training tasks, the training setting (e.g. computerized vs. paper-pencil vs. mixed, home-based vs. individual- vs. group settings), and its intensity. So far, however, the body of data is too small for subgroup analyses of subgroup analyses.

4.4 Strengths and Limitations

This systematic review is, to the authors' best knowledge, the first one systematically assessing prognostic factors and models for WMT responsiveness in healthy older adults on a single-person-within-study level rather than investigating moderating factors in a meta-analysis on a study-wide aggregated level as done in a recent meta-analysis on WMT in healthy older adults (Teixeira-Santos et al., 2019). Further strengths include the applied methods following the PICOTS system to define our review question, the CHARMS checklist for data extraction, and the PRISMA guidelines for the reporting of systematic reviews (Debray et al., 2017; Moher et al., 2009; Moons et al., 2014; Riley et al., 2019). One limitation is that due to insufficient reporting quality throughout many of the included studies, the studies in their entirety were sometimes difficult to comprehend, information might be misinterpreted by the reviewers, and results should be interpreted cautiously. Following, as already discussed above, due to methodological heterogeneity, we were not able to perform a quantitative meta-analysis, but had to focus on the qualitative directionality of the prognostic effects, limiting the validity of our findings. Furthermore, the applied WMT regimes within our included studies were highly heterogeneous regarding training duration, training tasks, and training setting. Only a multi-level IPD meta-analysis might be able to appropriately investigate the interplay of training-related and individual characteristics to answer the question "who benefits most". Additionally, the analyses to identify predictors of WMT responsiveness were conducted with data of the WMT groups only. Therefore, they did not control for effects in the control group (Hingorani et al., 2013), which impedes disentangling predictors of WMT responsiveness from predictors of retest and practice effects (Calamia, Markon, & Tranel, 2012). In this context, we need to admit that on the design stage of this systematic review, no comparator factor (C in PICOTS) was being considered, as our aim was to systematically assess any approach to prognostic research on WMT responsiveness. Furthermore, even though the risk of bias assessment followed the QUIPS checklist (Hayden et al., 2013) across six domains, the overall rating procedure across the items of one domain and across the six domains is not standardized by the developers.

4.5 Conclusion

Summarizing, prognostic research within the evaluation of WMT regimes in healthy older adults is still underrepresented given the urgent need for personalized cognitive prevention and intervention methods to counteract cognitive decline. Given the methodological shortcomings of the included studies, no clear conclusions can be drawn, and emerging patterns of prognostic effects will have to survive sound methodological replication in future attempts to promote precision medicine approaches in the context of WMT. However, within the small body of evidence and despite the complex relationships between cognitive reserve, neural plasticity and different proxies for these constructs, it seems that requirements for both, flexibility and plasticity have to be met. An IPD meta-analysis might be able to overcome the current research gaps regarding prognostic factors for WMT responsiveness in healthy older adults.

5 Conflict of Interest

AO reports no conflicts of interest.

MR has received a grant from the Brandau-Laibach Stiftung, and a grant from the German Ministry of Education and Research.

AKF has received a grant from the German Parkinson Society, and honoraria from ProLog Wissen GmbH, Cologne, Germany, pro audito Switzerland, Zürich, Switzerland, Seminar- und Fortbildungszentrum Rheine, Germany, and LOGOMANIA, Fendt & Sax GbR, Munich, Germany.

AFK is author of the cognitive training program NEUROvitalis but receives no corresponding honoraria.

NS reports no conflicts of interest.

EK has received grants from the German Ministry of Education and Research, ParkinsonFonds Deutschland gGmbH, the German Parkinson Society; honoraria from: Oticon GmbH, Hamburg, Germany; Lilly Pharma GmbH, Bad Homburg, Germany; Bernafon AG, Bern, Switzerland; Desitin GmbH, Hamburg, Germany. EK is author of the cognitive training program NEUROvitalis but receives no corresponding honoraria.

6 Author Contributions

AO, MR and EK conceptualized the presented work. MR conducted the systematic search, NS contributed to the systematic search. AO, MR, and AKF conducted the title and abstract screening. AO and MR conducted the full text screening, extracted the data, and conducted the risk of bias assessment. AO drafted the first version of the manuscript. All authors revised the manuscript for intellectual content and approved the final version of the manuscript. EK supervised the project during each stage of work.

7 Funding

This publication was funded by the Brandau-Laibach-Stiftung. No additional funding by funding agencies in the public, commercial, or not-for-profit sectors was obtained.

8 References

Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuehl, M., & Jaeggi, S. M. (2015). Improving fluid intelligence with training on working memory: a meta-analysis. *Psychonomic Bulletin & Review, 22*(2), 366-377.

Bäckman, L., & Nyberg, L. (2013). Dopamine and training-related working-memory improvement. *Neuroscience & Biobehavioral Reviews, 37*(9), 2209-2219.

Baltes, P. B., & Lindenberger, U. (1988). On the range of cognitive plasticity in old age as a function of experience: 15 years of intervention research. *Behavior therapy, 19*(3), 283-300.

Barulli, D., & Stern, Y. (2013). Efficiency, capacity, compensation, maintenance, plasticity: emerging concepts in cognitive reserve. *Trends in cognitive sciences, 17*(10), 502-509.

Beinhoff, U., Tumani, H., Brettschneider, J., Bittner, D., & Riepe, M. (2008). Gender-specificities in Alzheimer's disease and mild cognitive impairment. *Journal of Neurology, 255*(1), 117-122.

Bellander, M., Brehmer, Y., Westerberg, H., Karlsson, S., Fürth, D., Bergman, O., . . . Bäckman, L. (2011). Preliminary evidence that allelic variation in the LMX1A gene influences training-related working memory improvement. *Neuropsychologia, 49*(7), 1938-1942.

Berkowitz, C., Mosconi, L., Scheyer, O., Rahman, A., Hristov, H., & Isaacson, R. (2018). *Precision medicine for alzheimer's disease prevention.* Paper presented at the Healthcare.

Bewick, V., Cheek, L., & Ball, J. (2003). Statistics review 7: Correlation and regression. *Critical care, 7*(6), 451.

Borella, E., Carbone, E., Pastore, M., De Beni, R., & Carretti, B. (2017). Working memory training for healthy older adults: the role of individual characteristics in explaining short-and long-term gains. *Frontiers in human neuroscience, 11*, 99.

Borella, E., Carretti, B., Cantarella, A., Riboldi, F., Zavagnin, M., & De Beni, R. (2014). Benefits of training visuospatial working memory in young–old and old–old. *Developmental psychology, 50*(3), 714.

Borella, E., Carretti, B., Meneghetti, C., Carbone, E., Vincenzi, M., Madonna, J. C., . . . Mammarella, N. (2017). Is working memory training in older adults sensitive to music? *Psychological research, 83*(6), 1107-1123.

Borella, E., Carretti, B., Riboldi, F., & De Beni, R. (2010). Working memory training in older adults: evidence of transfer and maintenance effects. *Psychology and Aging, 25*(4), 767.

Borella, E., Carretti, B., Sciore, R., Capotosto, E., Taconnat, L., Cornoldi, C., & De Beni, R. (2017). Training working memory in older adults: Is there an advantage of using strategies? *Psychology and Aging, 32*(2), 178.

Borella, E., Carretti, B., Zanoni, G., Zavagnin, M., & De Beni, R. (2013). Working memory training in old age: An examination of transfer and maintenance effects. *Archives of Clinical Neuropsychology, 28*(4), 331-347.

Brehmer, Y., Rieckmann, A., Bellander, M., Westerberg, H., Fischer, H., & Bäckman, L. (2011). Neural correlates of training-related working-memory gains in old age. *Neuroimage, 58*(4), 1110-1120.

Brehmer, Y., Westerberg, H., Bellander, M., Fürth, D., Karlsson, S., & Bäckman, L. (2009). Working memory plasticity modulated by dopamine transporter genotype. *Neuroscience Letters, 467*(2), 117-120.

Brooks, J. O., Friedman, L., Pearman, A. M., Gray, C., & Yesavage, J. A. (1999). Mnemonic training in older adults: Effects of age, length of training, and type of cognitive pretraining. *International Psychogeriatrics, 11*(1), 75-84.

Brum, P. S., Borella, E., Carretti, B., & Sanches Yassuda, M. (2018). Verbal working memory training in older adults: an investigation of dose response. *Aging & Mental Health*, 1-11.

Burke, S. N., & Barnes, C. A. (2006). Neural plasticity in the ageing brain. *Nature Reviews Neuroscience, 7*(1), 30-40.

Bürki, C. N., Ludwig, C., Chicherio, C., & de Ribaupierre, A. (2014). Individual differences in cognitive plasticity: an investigation of training curves in younger and older adults. *Psychological research, 78*(6), 821-835.

Calamia, M., Markon, K., & Tranel, D. (2012). Scoring higher the second time around: meta-analyses of practice effects in neuropsychological assessment. *The Clinical Neuropsychologist, 26*(4), 543-570.

Carretti, B., Borella, E., Zavagnin, M., & de Beni, R. (2013). Gains in language comprehension relating to working memory training in healthy older adults. *International Journal of Geriatric Psychiatry, 28*(5), 539-546.

Chai, W. J., Abd Hamid, A. I., & Abdullah, J. M. (2018). Working Memory From the Psychological and Neurosciences Perspectives: A Review. *Frontiers in psychology, 9*, 401.

Chiu, H.-L., Chu, H., Tsai, J.-C., Liu, D., Chen, Y.-R., Yang, H.-L., & Chou, K.-R. (2017). The effect of cognitive-based training for the healthy older people: A meta-analysis of randomized controlled trials. *PloS one, 12*(5), e0176742.

Clark, D. O., Xu, H., Unverzagt, F. W., & Hendrie, H. (2016). Does targeted cognitive training reduce educational disparities in cognitive function among cognitively normal older adults? *International Journal of Geriatric Psychiatry, 31*(7), 809-817.

Colquitt, J. A., LePine, J. A., & Noe, R. A. (2000). Toward an integrative theory of training motivation: a meta-analytic path analysis of 20 years of research. *Journal of applied psychology, 85*(5), 678.

Dawson, N. V., & Weiss, R. (2012). Dichotomizing continuous variables in statistical analysis: a practice to avoid.

Deary, I. J., Corley, J., Gow, A. J., Harris, S. E., Houlihan, L. M., Marioni, R. E., . . . Starr, J. M. (2009). Age-associated cognitive decline. *British medical bulletin, 92*(1), 135-152.

Debray, T. P., Damen, J. A., Snell, K. I., Ensor, J., Hooft, L., Reitsma, J. B., . . . Moons, K. G. (2017). A guide to systematic review and meta-analysis of prediction model performance. *Bmj, 356*, i6460.

Diamond, A. (2007). Consequences of variations in genes that affect dopamine in prefrontal cortex. *Cerebral Cortex, 17*(suppl_1), i161-i170.

Double, K. S., & Birney, D. P. (2016). The effects of personality and metacognitive beliefs on cognitive training adherence and performance. *Personality and Individual Differences, 102*, 7-12.

Fairchild, J., Friedman, L., Rosen, A., & Yesavage, J. (2013). Which older adults maintain benefit from cognitive training? Use of signal detection methods to identify long-term treatment gains. *International Psychogeriatrics, 25*(4), 607-616.

Fernandes, A., Malaquias, C., Figueiredo, D., da Rocha, E., & Lins, R. (2019). Why Quantitative Variables Should Not Be Recoded as Categorical. *Journal of Applied Mathematics and Physics, 7*(07), 1519.

Frias, C. M. d., Annerbrink, K., Westberg, L., Eriksson, E., Adolfsson, R., & Nilsson, L.-G. (2005). Catechol O-methyltransferase Val158Met polymorphism is associated with cognitive performance in nondemented adults. *Journal of Cognitive Neuroscience, 17*(7), 1018-1025.

Greenwood, P. M., & Parasuraman, R. (2010). Neuronal and cognitive plasticity: a neurocognitive framework for ameliorating cognitive aging. *Frontiers in Aging Neuroscience, 2*, 150.

Hayden, J. A., van der Windt, D. A., Cartwright, J. L., Côté, P., & Bombardier, C. (2013). Assessing bias in studies of prognostic factors. *Annals of Internal Medicine, 158*(4), 280-286.

Heinzel, S., Lorenz, R. C., Brockhaus, W.-R., Wüstenberg, T., Kathmann, N., Heinz, A., & Rapp, M. A. (2014). Working memory load-dependent brain response predicts behavioral training gains in older adults. *Journal of Neuroscience, 34*(4), 1224-1233.

Heinzel, S., Riemer, T. G., Schulte, S., Onken, J., Heinz, A., & Rapp, M. A. (2014). Catechol-O-methyltransferase (COMT) genotype affects age-related changes in plasticity in working memory: a pilot study. *Biomed Research International, 2014*.

Hingorani, A. D., van der Windt, D. A., Riley, R. D., Abrams, K., Moons, K. G., Steyerberg, E. W., . . . Hemingway, H. (2013). Prognosis research strategy (PROGRESS) 4: stratified medicine research. *Bmj, 346*, e5793.

Hoenig, M. C., Bischof, G. N., Onur, Ö. A., Kukolja, J., Jessen, F., Fliessbach, K., . . . Drzezga, A. (2019). Level of education mitigates the impact of tau pathology on neuronal function. *European journal of nuclear medicine and molecular imaging, 46*(9), 1787-1795.

Kalbe, E., Bintener, C., Ophey, A., Reuter, C., Göbel, S., Klöters, S., . . . Kessler, J. (2018). Computerized Cognitive Training in Healthy Older Adults: Baseline Cognitive Level and Subjective Cognitive Concerns Predict Training Outcome. *Health, 10*(01), 20.

Li, S.-C. (2013). Neuromodulation and developmental contextual influences on neural and cognitive plasticity across the lifespan. *Neuroscience & Biobehavioral Reviews, 37*(9), 2201-2208.

López-Higes, R., Martín-Aragoneses, M. T., Rubio-Valdehita, S., Delgado-Losada, M. L., Montejo, P., Montenegro, M., . . . López-Sanz, D. (2018). Efficacy of cognitive training in older adults with and without subjective cognitive decline is associated with inhibition efficiency and working memory span, not with cognitive reserve. *Frontiers in Aging Neuroscience, 10*, 23.

Lövdén, M., Bäckman, L., Lindenberger, U., Schaefer, S., & Schmiedek, F. (2010). A theoretical framework for the study of adult cognitive plasticity. *Psychological Bulletin, 136*(4), 659.

913 Lövdén, M., Brehmer, Y., Li, S.-C., & Lindenberger, U. (2012). Training-induced compensation
914 versus magnification of individual differences in memory performance. *Frontiers in human*
915 *neuroscience, 6*, 141.
916 Lustig, C., Shah, P., Seidler, R., & Reuter-Lorenz, P. A. (2009). Aging, training, and the brain: a
917 review and future directions. *Neuropsychology Review, 19*(4), 504-522.
918 Matysiak, O., Kroemeke, A., & Brzezicka, A. (2019). Working memory capacity as a predictor of
919 cognitive training efficacy in the elderly population. *Frontiers in Aging Neuroscience, 11*.
920 McAvinue, L. P., Golemme, M., Castorina, M., Tatti, E., Pigni, F. M., Salomone, S., . . . Robertson,
921 I. H. (2013). An evaluation of a working memory training scheme in older adults. *Frontiers in*
922 *Aging Neuroscience, 5*, 20.
923 Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working memory training does not improve
924 performance on measures of intelligence or other measures of "far transfer" evidence from a
925 meta-analytic review. *Perspectives on Psychological Science, 11*(4), 512-534.
926 Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic
927 reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine, 151*(4),
928 264-269.
929 Mondini, S., Madella, I., Zangrossi, A., Bigolin, A., Tomasi, C., Michieletto, M., . . . Mapelli, D.
930 (2016). Cognitive reserve in dementia: implications for cognitive training. *Frontiers in Aging*
931 *Neuroscience, 8*, 84.
932 Moons, K. G., de Groot, J. A., Bouwmeester, W., Vergouwe, Y., Mallett, S., Altman, D. G., . . .
933 Collins, G. S. (2014). Critical appraisal and data extraction for systematic reviews of
934 prediction modelling studies: the CHARMS checklist. *Plos Medicine, 11*(10), e1001744.
935 Moons, K. G., Royston, P., Vergouwe, Y., Grobbee, D. E., & Altman, D. G. (2009). Prognosis and
936 prognostic research: what, why, and how? *Bmj, 338*, b375.
937 Moreau, D., Kirk, I. J., & Waldie, K. E. (2016). Seven pervasive statistical flaws in cognitive training
938 interventions. *Frontiers in human neuroscience, 10*, 153.
939 Newsom, J. T. (2015). *Longitudinal structural equation modeling: A comprehensive introduction*:
940 Routledge.
941 Noack, H., Lövdén, M., Schmiedek, F., & Lindenberger, U. (2009). Cognitive plasticity in adulthood
942 and old age: gauging the generality of cognitive intervention effects. *Restorative Neurology*
943 *and Neuroscience, 27*(5), 435-453.
944 Oswald, W. D., Gunzelmann, T., Rupprecht, R., & Hagen, B. (2006). Differential effects of single
945 versus combined cognitive and physical training with older adults: the SimA study in a 5-year
946 perspective. *European journal of ageing, 3*(4), 179.
947 Paraskevoudi, N., Balcı, F., & Vatakis, A. (2018). "Walking" through the sensory, cognitive, and
948 temporal degradations of healthy aging. *Annals of the New York Academy of Sciences,*
949 *1426*(1), 72-92.
950 Parish, A., Kim, J., Lewallen, K. M., Miller, S., Myers, J., Panepinto, R., & Maxwell, C. A. (2019).
951 Knowledge and perceptions about aging and frailty: An integrative review of the literature.
952 *Geriatric Nursing, 40*(1), 13-24.
953 Park, D. C., & Bischof, G. N. (2013). The aging mind: neuroplasticity in response to cognitive
954 training. *Dialogues in clinical neuroscience, 15*(1), 109.
955 Rahe, J., Liesk, J., Rosen, J. B., Petrelli, A., Kaesberg, S., Onur, O. A., . . . Kalbe, E. (2015). Sex
956 differences in cognitive training effects of patients with amnestic mild cognitive impairment.
957 *Aging, Neuropsychology, and Cognition, 22*(5), 620-638.
958 Reitz, C. (2016). Toward precision medicine in Alzheimer's disease. *Annals of translational*
959 *medicine, 4*(6).

Riley, R. D., Hayden, J. A., Steyerberg, E. W., Moons, K. G., Abrams, K., Kyzas, P. A., . . . Altman, D. G. (2013). Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. *Plos Medicine, 10*(2), e1001380.

Riley, R. D., Moons, K. G., Snell, K. I., Ensor, J., Hooft, L., Altman, D. G., . . . Debray, T. P. (2019). A guide to systematic review and meta-analysis of prognostic factor studies. *Bmj, 364*, k4597.

Roheger, M., Folkerts, A.-K., Krohm, F., Skoetz, N., & Kalbe, E. (2020). Prognostic factors for change in memory test performance after memory training in healthy older adults: A systematic review and outline of statistical challenges. *Diagnostic and Prognostic Research*.

Roheger, M., Meyer, J., Kessler, J., & Kalbe, E. (2019). Predicting short-and long-term cognitive training success in healthy older adults: who benefits? *Aging, Neuropsychology, and Cognition*, 1-19.

Sala, G., Aksayli, N. D., Tatlidil, K. S., Gondo, Y., & Gobet, F. (2019). Working memory training does not enhance older adults' cognitive skills: A comprehensive meta-analysis. *Intelligence, 77*, 101386.

Salthouse, T. A. (1990). Working memory as a processing resource in cognitive aging. *Developmental Review, 10*(1), 101-124.

Schubert, T., Strobach, T., & Karbach, J. (2014). New directions in cognitive training: on methods, transfer, and application. In: Springer.

Simon, S. S., Tusch, E. S., Feng, N. C., Hakansson, K., Mohammed, A. H., & Daffner, K. R. (2018). Is Computerized Working Memory Training Effective in Healthy Older Adults? Evidence from a Multi-Site, Randomized Controlled Trial. *Journal of Alzheimers Disease, 65*(3), 931-949. doi:10.3233/jad-180455

Smoleń, T., Jastrzebski, J., Estrada, E., & Chuderski, A. (2018). Most evidence for the compensation account of cognitive training is unreliable. *Memory & Cognition, 46*(8), 1315-1330.

Soveri, A., Antfolk, J., Karlsson, L., Salo, B., & Laine, M. (2017). Working memory training revisited: A multi-level meta-analysis of n-back training studies. *Psychonomic Bulletin & Review, 24*(4), 1077-1096.

Stern, Y. (2002). What is cognitive reserve? Theory and research application of the reserve concept. *Journal of the International Neuropsychological Society, 8*(3), 448-460.

Stern, Y. (2009). Cognitive reserve. *Neuropsychologia, 47*(10), 2015-2028.

Stern, Y., Arenaza-Urquijo, E. M., Bartrés-Faz, D., Belleville, S., Cantilon, M., Chetelat, G., . . . Kremen, W. S. (2018). Whitepaper: Defining and investigating cognitive reserve, brain reserve, and brain maintenance. *Alzheimer's & Dementia*.

Stern, Y., Barnes, C. A., Grady, C., Jones, R. N., & Raz, N. (2019). Brain reserve, cognitive reserve, compensation, and maintenance: operationalization, validity, and mechanisms of cognitive resilience. *Neurobiology of Aging, 83*, 124-129.

Studer-Luethi, B., Jaeggi, S. M., Buschkuehl, M., & Perrig, W. J. (2012). Influence of neuroticism and conscientiousness on working memory training outcome. *Personality and Individual Differences, 53*(1), 44-49.

Teixeira-Santos, A. C., Moreira, C. S., Magalhães, R., Magalhães, C., Pereira, D. R., Leite, J., . . . Sampaio, A. (2019). Reviewing working memory training gains in healthy older adults: A meta-analytic review of transfer for cognitive outcomes. *Neuroscience & Biobehavioral Reviews*.

Tusch, E. S., Alperin, B. R., Ryan, E., Holcomb, P. J., Mohammed, A. H., & Daffner, K. R. (2016). Changes in neural activity underlying working memory after computerized cognitive training in older adults. *Frontiers in Aging Neuroscience, 8*, 255.

Valenzuela, M. J., & Sachdev, P. (2006). Brain reserve and cognitive decline: a non-parametric systematic review. *Psychological Medicine, 36*(8), 1065-1073.

Verhaeghen, P., Marcoen, A., & Goossens, L. (1992). Improving memory performance in the aged through mnemonic training: a meta-analytic study. *Psychology and Aging, 7*(2), 242.
von Bastian, C. C., & Eschen, A. (2016). Does working memory training have to be adaptive? *Psychological research, 80*(2), 181-194.
Weicker, J., Hudl, N., Frisch, S., Lepsien, J., Mueller, K., Villringer, A., & Thoene-Otto, A. (2018). WOME: Theory-based working memory training—A placebo-controlled, double-blind evaluation in older adults. *Frontiers in Aging Neuroscience, 10*, 247.
Weicker, J., Villringer, A., & Thöne-Otto, A. (2016). Can impaired working memory functioning be improved by training? A meta-analysis with a special focus on brain injured patients. *Neuropsychology, 30*(2), 190-212.
West, R. L., Bagwell, D. K., & Dark-Freudeman, A. (2008). Self-efficacy and memory aging: The impact of a memory intervention based on self-efficacy. *Aging, Neuropsychology, and Cognition, 15*(3), 302-329.
Whitlock, L. A., McLaughlin, A. C., & Allaire, J. C. (2012). Individual differences in response to cognitive training: Using a multi-modal, attentionally demanding game-based intervention for older adults. *Computers in human behavior, 28*(4), 1091-1096.
Willis, S. L., & Caskie, G. I. (2013). Reasoning training in the ACTIVE study: how much is needed and who benefits? *Journal of Aging and Health, 25*(8_suppl), 43S-64S.
Wilson, R. S., Boyle, P. A., Yu, L., Barnes, L. L., Schneider, J. A., & Bennett, D. A. (2013). Life-span cognitive activity, neuropathologic burden, and cognitive aging. *Neurology, 81*(4), 314-321.
Wolff, R. F., Moons, K. G., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., . . . Mallett, S. (2019). PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Annals of Internal Medicine, 170*(1), 51-58.
Zinke, K., Zeintl, M., Eschen, A., Herzog, C., & Kliegel, M. (2012). Potentials and limits of plasticity induced by working memory training in old-old age. *Gerontology, 58*(1), 79-87.
Zinke, K., Zeintl, M., Rose, N. S., Putzmann, J., Pydde, A., & Kliegel, M. (2014). Working memory training and transfer in older adults: effects of age, baseline performance, and training gains. *Developmental psychology, 50*(1), 304.

9 Data Availability Statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Table 1. Study Objectives, Participants' Demographics and Working Memory Training Characteristics

Study	**Analysis**		**Prognostic Factors**		**Participants**					**Training**	
Author (year)	**Prognostic Model**	**Factor Finding**	**Individual-related**	**Training-related**	**n**[a]	**Age** (in years)	**Sex**	**Education** (in years)	**Global Cognition**	**Total Time of Training** (in minutes) **and Setting**	**Description of Training**
Borella et al. (2013)		X	X		38	young-old[b] 69.00 (3.18), 65-75; old-old 79.22 (3.49), 75-87	young-old 13 ♀, 7 ♂; old-old 12 ♀, 6 ♂	young-old 9.40 (3.95); old-old 5.72 (2.52)	vocabulary score WAIS-R[c], max. 70; young-old 46.65 (8.64); old-old 42.72 (9.04)	180 (3 sessions of 60 minutes over 2 weeks) individual setting	adaptive verbal working memory training with the Categorization Working Memory Span (CWMS[d]) Task via audio-recordings
Borella et al. (2014)		X	X	X	40 40	young-old 69.90 (2.79), 65-75; old-old 79.60 (2.28), 76-84	n.a.	young-old 10.65 (2.50); old-old 8.75 (1.33)	vocabulary score WAIS-R[c], max. 70; young-old 49.25 (5.82); old-old 50.15 (4.57)	180 (3 sessions of 60 minutes over 2 weeks) individual setting	adaptive visuospatial working memory training with a computerized version of the Matrix Task[e]
Borella, Carbone, et al. (2017)[f]	X		X		73	71.63 (5.53), 61-87	n.a.	9.42 (4.54)	vocabulary score WAIS-R[c], max. 70; 49.21 (10.89)	180 (3 sessions of 60 minutes over 2 weeks) individual setting	adaptive verbal working memory training with the CWMS[d] Task via audio-recordings
Borella, Carretti, Meneghetti, et al. (2017)		X		X	54	Mozart 70.15 (2.79); Albinoni 69.31 (3.30); White Noise 68.18 (3.48); 65-75	Mozart 11 ♀, 8 ♂; Albinoni 7 ♀, 12 ♂; White Noise 12 ♀, 4 ♂	Mozart 13.84 (2.91); Albinoni 14.73 (2.15); White Noise 13.06 (4.00)	n.a.	180 (3 sessions of 60 minutes over 2 weeks) individual setting	6 minutes of listening to music according to experimental condition followed by adaptive verbal working memory training with the CWMS[d] Task via audio-recordings
Borella, Carretti, Sciore, et al. (2017)	X		X	X	36	WM 69.44 (3.73); WM+S 67.94 (4.89)	WM 10 ♀, 8 ♂; WM+S 13 ♀, 5 ♂	WM 14.39 (2.87); WM+S 13.56 (2.92)	vocabulary score WAIS-R[c], max. 70; WM 61.72 (5.63); WM+S 58.39 (9.89)	105 (3 sessions of 35 minutes over 2 weeks) individual setting	adaptive verbal working memory training with the CWMS[d] Task via audio-recordings; for the WM+S group preliminary instructions to use a visual mental imagery strategy[g]
Brehmer et al. (2011)		X	X	X	24	63.6 (SD n.a.); 60-70	12 ♀, 12 ♂	n.a.	n.a.	625 (25 sessions of 25 minutes over 5 weeks) home-based individual setting	adaptive vs. non-adaptive (fixed at low level) both verbal and visuospatial working memory training with the computerized Cogmed[h] training program
Brum et al. (2018)		X		X	41	3 sessions 67.17 (4.40); 6 sessions 67.91 (3.61)	n.a.	3 sessions 9.50 (5.25); 6 sessions 7.57 (3.34)	Clock Drawing Test[i], max. 10; 3 sessions 9.00 (1.13); 6 sessions 8.83 (0.98)	3 sessions: 105 (3 sessions of 35 minutes over 1 week) 6 sessions: 210 (6 sessions of 35 minutes over 2 weeks) individual setting	adaptive verbal working memory training with the CWMS[d] Task via audio-recordings
Heinzel, Lorenz, et al. (2014)	X	X	X		19	66.0 (3.73); 61-75	6 ♀, 13 ♂	15.61 (3.26)	n.a.	540 (12 sessions of 45 minutes over 4 weeks) individual setting	adaptive computerized numerical n-back training paradigm[l]

Heinzel, Riemer, et al. (2014)		X	X		25	Val/Val 67.36 (4.34); any Met 64.64 (3.37)	Val/Val 5 ♀, 6 ♂; any Met 7 ♀, 7 ♂	Val/Val 15.46 (3.15); any Met 16.88 (3.62)	MMSE[k], max. 30; Val/Val 29.27 (1.01); any Met 29.64 (0.84)	540 (12 sessions of 45 minutes over 4 weeks) individual setting	adaptive computerized numerical n-back training paradigm[l]
Matysiak et al. (2019)		X	X		43	65.9 (SD n.a.)	28 ♀, 15 ♂	n.a. education dichotomized into higher (n=27) vs. secondary (n=16) education	operation span (OSPAN)[m] score, max. n.a.; 15.31 (SD n.a.)	500 (25 sessions of 20 minutes over 5 weeks) home-based individual setting	adaptive computerized dual (visuo-spatial and auditory/verbal) n-back training paradigm[n]

McAvinue et al. (2013)		X		X	19	69.89 (4.5); 64-79	13 ♀, 6 ♂	n.a.
								educational levels only:
								primary school n=1;
								leaving certificate n=2;
								undergraduate n=10;
								postgraduate n=6
									MMSE[k], max. 30;
									27.74 (2.05);
									AMNART IQ[o], max. n.a.;
									120.47 (4.44)
										750 (25 sessions of 30 minutes over 5 weeks)
										home-based individual setting
											adaptive computerized verbal and visuo-spatial working memory training plus psychoeducation on everyday cognitive strategies

Simon et al. (2018)	X		X	X	82	adaptive 72.4 (5.6); non-adaptive 73.7 (6.5)	adaptive 29 ♀, 12 ♂; non-adaptive 26 ♀, 15 ♂	adaptive 15.7 (3.7); non-adaptive 15.3 (3.2)	MMSE[k], max. 30; adaptive 29.2 (1.1); non-adaptive 29.0 (1.3); AMNART IQ[o], max. n.a.; adaptive 122.6 (5.9); non-adaptive 120.6 (6.0)	1000 (25 sessions of 40 minutes over 5 weeks) home-based individual setting	adaptive vs. non-adaptive (fixed at low level) both verbal and visuospatial working memory training with the computerized Cogmed[h] training program
Tusch et al. (2016)		X	X	X	41	adaptive 74.47 (6.26); non-adaptive 76.84 (5.95)	adaptive 12 ♀, 5 ♂; non-adaptive 15 ♀, 3 ♂	adaptive 18.65(2.98); non-adaptive 16.78(2.05)	MMSE[k], max. 30; adaptive 29.41 (0.71); non-adaptive 28.89 (1.68); AMNART IQ[o], max. n.a.; adaptive 123.59 (4.00); non-adaptive 119.33 (5.86)	1000 (25 sessions of 40 minutes over 5 weeks) home-based individual setting	adaptive vs. non-adaptive (fixed at low level) both verbal and visuospatial working memory training with the computerized Cogmed[h] training program
Weicker et al. (2018)		X	X	X	40	adaptive 67.8 (3.9); non-adaptive 67.7 (3.1)	adaptive 10 ♀, 10 ♂; non-adaptive 11 ♀, 9 ♂	n.a. categorized only: <9 years n=4; 10-12 years n=16; >12 years n=20	n.a.	540 (12 sessions of 45 minutes over 4 weeks) individual setting	adaptive vs. non-adaptive (fixed at low-level) working memory training with the computerized WOME[q] (WOrking MEmory) training program
Zinke et al. (2012)		X	X		20	86.8 (4.9); 77-96	14 ♀, 6 ♂	11.7 (3.3)	MMST short form for old-old adults[r], max. 21; 19.4 (1.4)	275 (10 sessions of 25-30 minutes over 2 weeks) individual setting	adaptive paper-pencil verbal and visuo-spatial working memory training
Zinke et al. (2014)	X		X		40	76.7 (8.4); 65-95	32 ♀, 8 ♂	14.4 (3.4)	MMST short form for old-old adults[r], max. 21; 20.2 (1.1)	270 (9 sessions of 30 minutes over 3 weeks) individual setting	adaptive paper-pencil verbal and visuo-spatial working memory and executive control training

Note.

[a] number of participants in the working memory training group

[b] young-old sample from Borella, E., Carretti, B., Riboldi, F., & De Beni, R. (2010). Working memory training in older adults: evidence of transfer and maintenance effects. Psychology and aging, 25(4), 767.

[c] WAIS-R, Wechsler Adult Intelligence Scale-revised manual. Wechsler, D. (1981). Wechsler Adult Intelligence Scale-revised manual. New York, NY: Psychological Corporation.

[d] CWMS, Categorization Working Memory Span. De Beni, R., Borella, E., Carretti, B., Marigo, C., and Nava, L. (2008). BAC. Portfolio per la Valutazione del Benessere e delle Abilità Cognitive nell'età adulta e Avanzata [The Assessment of Well-being and Cognitive Abilities in Adulthood and Aging]. Firenze: Giunti, OS. Training procedure introduced by Borella et al. (2010)

[e] adapted from Cornoldi, C., Bassani, C., Berto, R., & Mammarella, N. (2007). Aging and the intrusion superiority effect in visuo-spatial working memory. Aging, Neuropsychology, and Cognition, 14, 1–21. and Carretti, B., Mammarella, I. C., & Borella, E. (2012). Age differences in proactive interference in verbal and visuo-spatial working memory. Journal of Cognitive Psychology, 24, 243–255.

[f] post-hoc analysis of: Borella et al. (2010); Borella et al. (2013); Borella 2017 (3); Carretti, B., Borella, E., Zavagnin, M., and De Beni, R. (2013). Gains in language comprehension relating to working memory training in healthy older adults. Int. J. Geriatr. Psychiatry 28, 539–546

[g] as described in Carretti, B., Borella, E., & De Beni, R. (2007). Does strategic memory training improve the working memory performance of younger and older adults? Experimental Psychology, 54, 311–320.

[h] for details about the adaptive training algorithm, see Cogmed QM; www.cogmed.com; Klingberg, T., Forssberg, H., Westerberg, H., 2002. Increased brain activity in frontal and parietal cortex underlies the development of visuospatial working memory capacity during childhood. J. Cogn. Neurosci. 14, 1–10.

[i] Aprahamian, I., Martinelli, J. E., Neri, A. L., & Yassuda, M. S. (2010). The accuracy of the Clock Drawing Test compared to that of standard screening tests for Alzheimer's disease: Results from a study of Brazilian elderly with heterogeneous educational backgrounds. International Psychogeriatrics, 22(01), 64–71.Shulman, K. I. (2000). Clock-drawing: Is it the ideal cognitive screening test?. International Journal of Geriatric Psychiatry, 15(6), 548–561.

[j] Carriers of either Val/Met or Met/Met COMT (catechol-O-methyltransferase) genotype were classified into one group (any Met) and contrasted with Val/Val carriers

[k] MMSE, Mini-Mental State Examination; Folstein, M.F., Folstein, S.E., White, T., and Messer, M.A. (2010). Mini Mental State Examination, 2nd Edn. Lutz: Psychological Assessment Resources, Inc.

[l] Gevins, A., & Cutillo, B. (1993). Spatiotemporal dynamics of component processes in human working memory. *Electroencephalography and clinical Neurophysiology*, *87*(3), 128-143.

[m] computerized version of the original OSPAN task; Turner, M. L., and Engle, R. W. (1989). Is working memory capacity task dependent? J. Mem. Lang. 28, 127–154.

[n] introduced by Jaeggi, S. M., Buschkuehl, M., Jonides, J., and Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. PNAS 105, 6829–6833.

[o] AMNART, American National Adult Reading Test; Nelson, H. (1982).National Adult Reading Test Manual. Windsor: NFER-Nelson.

[p] TMT-A, Trail Making Test part A; Reitan R, Wolfson D (1985) The Halstead-Reitan Neuropsycholgical Test Battery: Therapy and clinical interpretation. Neuropsychological Press, Tucson, AZ.

[q] WOME, WOrking MEmory; part of the cognitive rehabilitation program RehaCom®

[r] Mini-Mental State Examination short form for old-old adults by Kliegel, M., Rott, C., d'Heureuse, V., Becker, G., Schönemann, P. (2001) Demenz im höchsten Alter ist keine Notwendigkeit. Ergebnisse der Heidelberger Hundertjährigen-Studie. Z Gerontopsychol Psychiatr; 14: 169–180.

[s] Lehrl, S. (2005). Mehrfachwahl-Wortschatz-Intelligenztest (MWT–B) [Multiple-choice vocabulary–intelligence test]. Balingen, Germany: Spitta.

Table 2. Risk of Bias Assessment Using the QUIPS Checklist

Author (year)	Study Participation	Study Attrition	Prognostic Factor Measurement	Outcome Measurement	Study Confounding	Statistical Analysis and Reporting
Borella et al. (2013)						
Borella et al. (2014)						
Borella, Carbone, et al. (2017)						
Borella, Carretti, Meneghetti, et al. (2017)						
Borella, Carretti, Sciore, et al. (2017)						
Brehmer et al. (2011)						
Brum et al. (2018)						
Heinzel, Lorenz, et al. (2014)						
Heinzel, Riemer, et al. (2014)						
Matysiak et al. (2019)						
McAvinue et al. (2013)						
Simon et al. (2018)						
Tusch et al. (2016)						
Weicker et al. (2018)						
Zinke et al. (2012)						
Zinke et al. (2014)						

Note. Overall risk of bias rating of domains in the Quality in Prognosis Studies (QUIPS) checklist (Hayden et al. 2013). Red = high risk, yellow = moderate risk, green = low risk. For details on individual items and rating scheme, please refer to Supplementary Material 3.

Table 3. Prognostic Analyses, Outcomes, Results, and Timing

Study	Analysis				Outcome	Prediction Results									Timing	
	Model	Factor Finding				Degree of transfer	baseline performance	intelligence	age	education	sex	adaptivity	dose of training	others	Post-Intervention	Follow-Up
		Corr/Reg	GLM	Others												
Borella et al. (2013)				X effect size	verbal working memory Δd	+			↓						X	X 8
						†			↓						X	
					visuospatial working memory Δd	†			↓						X	
					short-term memory Δd	‡			↓						X	
					fluid intelligence Δd	‡			↓						X	
					processing speed Δd	‡			↓						X	
					inhibition Δd	‡			↓						X	
						‡			↓						X	
Borella et al. (2014)			X ANOVA		verbal working memory Δs	+			↓						X	X 8
					visuospatial working memory Δs	†			↓						X	X 8
Borella et al. (2014)				X effect size	working memory Δd	†								training modality: visuospatial --	--	
					verbal working memory Δd	+								training modality: visuospatial --	--	
					visuospatial working memory Δd	†								training modality: visuospatial --	--	
					short-term memory Δd	‡								training modality: visuospatial --	--	
					fluid intelligence Δd	‡								training modality: visuospatial ↓	--	
					processing speed Δd	‡								training modality: visuospatial --	--	
					inhibition Δd	‡								training modality: visuospatial ↓	--	
Borella, Carbone, et al. (2017)	X linear mixed models				verbal working memory	+		↓							X	X 8
						†			↓	--					X	-- 8
					visuospatial working memory	†		↑	↓						X	-- 8
					short-term memory	‡	↓	↓		↓					X	-- 8
					fluid intelligence	‡	↓		↓						X	X 8
					processing speed	‡	--	↑							X	-- 8
					inhibition	‡			↓						X	-- 8
Borella, Carretti, Meneghetti, et al. (2017)			X ANOVA		verbal working memory	+								music listening condition: Albinoni ↑	X	-- 6
					visuospatial working memory	†								music listening condition --	--	-- 6
					fluid intelligence	‡								music listening condition: Albinoni ↑	X	-- 6
					phonemic verbal fluency	‡								music listening condition --	--	-- 6
	X				verbal working memory Δ	+	↓							strategy use ↑	X	

Borella, Carretti, Sciore, et al. (2017)	hierarchical regression					†	↓												X 8
						†	↓												X 8
						†	↓												X 8
					visuospatial working memory Δ	†	↓												X 8
					processing speed Δ	‡	↓											X	
						‡	↓												X 8
Brehmer et al. (2011)			X ANOVA		verbal working memory	†					--						--		
					visuospatial working memory	†					↑						X		
					short-term memory	‡					--						--		
						‡					--						--		
					episodic memory	‡					↑						X		
					attention	‡					↑						X		
					reasoning	‡					--						--		
					inhibition	‡					--						--		
		X Pearsson			working memory Δmax	+	↑*										X		
Brum et al. (2018)			X ANOVA		verbal working memory	+						--					--		-- 6
						†						--					--		-- 6
						†						--					--		-- 6
					visuospatial working memory	†						--					--		-- 6
						†						--					--		-- 6
					verbal short-term memory	‡						--					--		-- 6
					visuospatial short-term memory	‡						--					--		-- 6
					reasoning	‡						--					--		-- 6
					inhibition	‡						--					--		-- 6
					semantic fluency	‡						--					--		-- 6
				X effect size	verbal working memory Δd	+						--					--		-- 6
						†						↑					--		X 6
						†						↓					X		X 6
					visuospatial working memory Δd	†						↓					--		X 6
						†						--					--		-- 6
					verbal short-term memory Δd	‡						--					--		-- 6
					visuospatial short-term memory Δd	‡						↓					--		X 6
					reasoning Δd	‡						↑					X		X 6
					inhibition Δd	‡						↑					X		X 6
					semantic fluency Δd	‡						↓					--		X 6
(Heinzel, Lorenz, et al., 2014)	X hierarchical regression				verbal working memory Δ	+			↑	↓	?					baseline load-dependent BOLD ↑ gray matter volume ↑	X		
						+	↑		?	?	?					baseline load-dependent BOLD ↑ gray matter volume?	X		
		X			verbal working memory Δ	+										baseline load-dependent BOLD ↑	X		

		Pearsson														
Heinzel, Riemer, et al. (2014)			X ANOVA		verbal working memory	+								Val/Val ↓	X	
Matysiak et al. (2019)		X linear mixed models			verbal working memory max	+	↑*									
						+			--							
						+				--						
						+					--					
						+								occupational activity --		
McAvinue et al. (2013)		X Pearsson			short-term memory	‡							↓		X	
						‡							--		--	
					long-term memory	‡							--		--	
					anxiety & depression	‡							↓		X	
Simon et al. (2018)	X linear mixed models				working memory	†						↑		processing speed ?	X	
						†			--			↑			X	
					processing speed	‡			--			--			--	
					set shifting	‡			--			--			--	
					phonemic fluency	‡			--			--			--	
					semantic fluency	‡			--			--			--	
Tusch et al. (2016)			X ANOVA		verbal working memory	†						--			--	
		X Pearsson			verbal working memory Δ	†		--							--	
						†				--					--	
Weicker et al. (2018)			X ANOVA		working memory	†						--			--	-- 3
					working memory span	†						↑			X	-- 3
					visuospatial working memory	†						↑			X	-- 3
					executive functions	‡						--			--	-- 3
					logical reasoning	‡						--			--	-- 3
					long-term memory	‡						--			--	-- 3
		X Pearsson			working memory Δmax	+	↑*								X	
						+	↑*								X	
						+	↑*								X	
Zinke et al. (2012)			X t-tests		working memory Δ	+	↓								X	
		X Pearsson			verbal working memory Δ	†	↓								X	
					visuospatial working memory Δ	†	↓								X	
						†	↓								X	
					verbal short-term memory Δ	‡	↓								X	
					visuospatial short-term memory Δ	‡	↓								X	

Zinke et al. (2014)	X hierarchical regression		verbal working memory Δ	+	↓	--	↓				X
				†		--	--			training task gains ↑	X
			visuospatial working memory Δ	+	↓	--	↓				X
				†		--	↓			training task gains --	X
			executive control Δ	+	↓	↑	--				X
				†		--	↓			training task gains ↓↑	X
			fluid intelligence Δ	‡		--	↑			training task gains --	X
			Inhibition Δ	‡		↑	↓			training task gains --	X

Note. Detailed information on the analytical plan of each prediction approach, the operationalization of both prognostic variables and outcomes, and extracted results can be obtained from Supplementary Material 4 and 5. If not indicated otherwise, time was included as one factor in the analyses, therefore abandoning the use of change scores as dependent variable and investigating more than one point of time (e.g. pre-test, post-test, and follow-up) within one analysis. Within prediction results, ↑ indicates positive predictors, i.e. higher values in the predictor variable are associated with better training outcomes, ↓ indicates negative predictors, i.e. lower values in the predictor variable are associated with better training outcomes, -- indicates non-significant relationships between predictor and training outcome, and ? indicates that a predictor was investigated, but results were not reported appropriately. For prognostic model studies, only the final models reported in the original manuscript are reported. In the Timing column, an X indicates a significant influence of the prognostic factor(s) under investigation on the respective outcome at the given point of time, -- non-significant relationships only. For follow-ups, time in months is indicated.

Δ, unstandardized / raw change score as dependent variable; Δd, Cohen's d as dependent variable; Δmax, maximum change score; Δs, standardized change score as dependent variable ((post-pre)/SD_{pre}); Corr/Reg, analytical approaches including correlations, linear regressions, multi-level modeling approaches; GLM, Generalized Linear Model approaches including ANOVA, ANCOVA, independent sample t-tests

* dependent variable represents the maximum level / change achieved during training

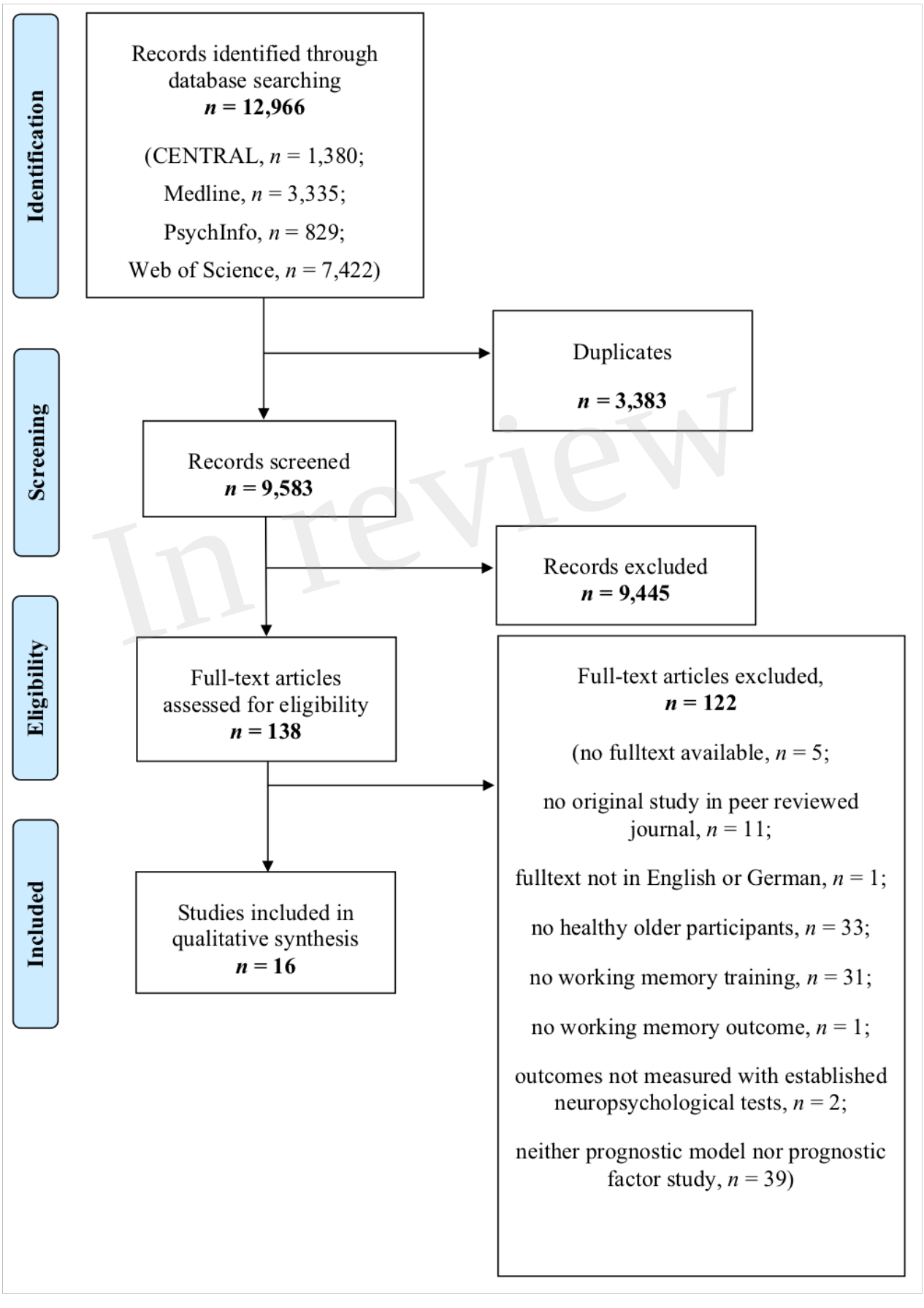
+ direct training effect, i.e. task was trained within the working memory training

† near-transfer effect, i.e. task was not trained within the working memory training, but represents (verbal and/or visuospatial) working memory

‡ far-transfer effect, i.e. task was not trained within the working memory training and does not represent (verbal and/or visuospatial) working memory

Figure 1. PRISMA Flow Diagram

Figure 1.TIFF



Prognostic models for changes in memory performance after memory training in healthy older adults: A systematic review

Short Title: Prognostic Models for Memory Changes

Mandy Roheger [a,b], Ann-Kristin Folkerts[a], Fabian Krohm [a], Nicole Skoetz[c] & Elke Kalbe[a]

[a] *Department of Medical Psychology | Neuropsychology and Gender studies & Center for Neuropsychological Diagnostics and Intervention (CeNDI), Faculty of Medicine and University Hospital Cologne, University of Cologne, Kerpener Str. 68, 50937 Cologne, Germany*

[b] *Department of Neurology, University Medicine Greifswald, Greifswald, Germany*

[c]*Evidence-Based Oncology, Department I of Internal Medicine, Faculty of Medicine and University Hospital Cologne, University of Cologne, Kerpener Str. 62, 50924 Cologne, Germany*

Correspondence to: Elke Kalbe, Department of Medical Psychology | Neuropsychology and Gender Studies & Center for Neuropsychological Diagnostics and Intervention (CeNDI), Faculty of Medicine and University Hospital Cologne, University of Cologne Kerpener Str. 68, 50937 Köln, Germany, phone +49 221 478-6669, fax +49 221 478-3420, elke,kalbe@uk-koeln.de

FUNDING

The study was funded by the Brandau-Laibach foundation of the University Hospital of Cologne, Germany.

CONFLICTS OF INTEREST

MR has received a grant from the Brandau-Laibach Stiftung, and a grant from the German Ministry of Education and Research.

AKF has received a grant from the German Parkinson Society, and honoraria from ProLog Wissen GmbH, Cologne, Germany, pro audito Switzerland, Zürich, Switzerland, LOGOMANIA, Fendt & Sax GbR, Muenchen, Germany, and Seminar- und Fortbildungszentrum Rheine, Germany. AFK is author of the cognitive training program NEUROvitalis but receives no corresponding honoraria.

FK and NS do not declare any conflict of interests.

EK has received grants from the German Ministry of Education and Research, ParkinsonFonds Deutschland GmbH, the German Parkinson Society; honoraria from: Oticon GmbH, Hamburg, Germany; Lilly Pharma GmbH, Bad Homburg, Germany; Bernafon AG, Bern, Switzerland; Desitin GmbH, Hamburg, Germany. EK is author of the cognitive training program NEUROvitalis but receives no corresponding honoraria.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Abstract

Background: Identifying individuals' profiles of prognostic factors that predict improvements after nonpharmacological interventions such as memory trainings may help to not only predict individuals' future outcome after such intervention, but also to tailor new trainings for individuals with specific characteristics. However, until now, no systematic review on prognostic models, defined as a set of multiple prognostic factors to predict a future outcome, for changes in memory performance after memory training exist.

Methods: MEDLINE, Web of Science Core Collection, CENTRAL, and PsycInfo were searched up to November 2019 to identify studies investigating prognostic models on verbal and non-verbal short- and long-term memory after conducting memory training in healthy older adults. The PROBAST tool was used to assess Risk of Bias.

Results: After screening *n* = 10,703 studies, *n* = 12 studies were included. These studies and the investigated statistical models are highly heterogeneous, so that conclusions are limited. However, one consistent result was that lower age combined with higher education seems to predict improvements after memory training.

Conclusion: More studies on prognostic models for memory changes after memory training have to be conducted before clear conclusions which will help to tailor memory trainings to individuals' profiles can be drawn.

Registration: CRD42018105803, https://www.crd.york.ac.uk/PROSPERO

Keywords: prognostic model, memory training, cognition, memory, healthy aging

Background

A prognostic or predictive model is a formal combination of multiple predictors from which risks of a specific endpoint can be calculated for individuals (Steyerberg et al., 2013). Prognostic models are regularly used in medical research; however, their use in neuropsychological research to predict changes after nonpharmacological interventions, e.g. cognitive training (CT) is rather limited. As data demonstrates that CT (i.e., a structured approach to strengthen targeted cognitive functions, e.g. memory, attention and executive functions with the help of specific paper and pencil or cognitive tasks) is effective in improving cognitive outcomes in healthy older adults (Chiu et al., 2017), identifying individuals' profiles of prognostic factors that predict improvements after these kind of interventions may help to not only predict individuals' future outcome after CT. Further, it may improve informed decision making among clinicians to follow a personalized medicine approach (Altman, Vergouwe, Royston, & Moons, 2009), and it can also be used to improve the design and analysis of randomised therapeutic trials while considering person-centered intervention programs (Roozenbeek et al., 2009).

One particular form of CT targets memory functions and/or the use of memory strategies. Memory decline is a common process among older adults and may affect their ability to function independently in our society (Verhaeghen, Geraerts, & Marcoen, 2000). Also, pathological memory impairment is indicative of neurodegenerative diseases such as dementia (Jockwitz et al., 2019). Yet, memory training is an effective method for modifying not only trained memory function, but also maintaining further non-trained memory functions as well as non-cognitive abilities in older adults (Hitchcock, Werner-Seidler, Blackwell, & Dalgleish, 2017; Rosi et al., 2018; Simon et al., 2018). Notably, results from the literature indicates that there is a great variability of responsiveness among healthy older training participants (Langbaum, Rebok, Bandeen-Roche, & Carlson, 2009) on the first sight. A recently published systematic review on prognostic factors on memory changes after memory

training in healthy older adults showed high between-study heterogeneity with regard to the assessment, statistical evaluation, and reporting of the investigated prognostic factors. Included studies used different types of dependent variables (change scores vs. post-test scores) when defining memory training success leading to contradictory results. Age was the only variable investigated throughout most of the studies, showing that older adults benefit more from training when using the change score as the dependent variable. Further, the review could show that the tendency of the prognostic factor (the more of x/the less of x versus the more of x/the less of y) is dependent on the used dependent outcome measure of the studies (e.g., whether post-test scores or changes scores were used in calculations as the dependent variable, Roheger, Folkerts, Krohm, Skoetz, & Kalbe, 2020). Yet, this review focused on prognostic factors, defined as any measure that, among people with a given condition (process of aging, the start point), is associated with a subsequent outcome (an endpoint, worsening of cognition, Riley et al., 2013). Until now, no systematic review investigates prognostic models for changes in memory outcomes after conducting memory training. Prognostic models are defined as a set of multiple prognostic factors to predict a future outcome. Yet, prognostic models take into account multiple factors and their variances, with the ability to reveal potential suppressing factors, provide different information than prognostic factor studies, and have to be assessed with different tools regarding risk of bias judgement. Therefore, the present paper systematically summarizes prognostic models of memory changes after memory training in healthy older adults (≥ 55 years) and discusses different statistical methods used to calculate prognostic models.

METHODS

The reporting of the present review follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guideline (Moher, Liberati, Tetzlaff, & Altman, 2009). The “PRISMA for Abstracts Checklist” and the “PRISMA checklist for

systematic reviews" are depicted in Supplementary Tables 1 and 2. The pre-registered review protocol can be assessed at [blinded for peer review].

Search and study selection

MEDLINE Ovid, Web of Science Core Collection, CENTRAL and PsycInfo were systematically searched up to October 2018. An update-search was conducted until November 2019. Further, reference lists of all identified trials, relevant review articles and current treatment guidelines were hand searched. If no full text could be obtained, the authors were contacted and asked to provide full text publications within a two-week time frame. The full search strings for each database are presented in the Supplementary Material, Table 3 – 6.

Two review authors ([blinded for peer review]) screened titles and abstracts according to the predefined eligibility criteria. Full-text articles, whose abstracts met the inclusion criteria, were further reviewed by two authors ([blinded for peer review]) for inclusion in the review. In cases where no consensus could be reached, a third author ([blinded for peer review]) was asked and the case was discussed until a final consensus was obtained.

Eligibility criteria

The review focused on peer-reviewed studies with no limitations regarding publication date which investigated prognostic models of changes in memory test performance after memory training. The studies could be published in English or German. Full study reports needed to be available. We excluded abstracts, books, book chapters, study protocols, and conference abstracts. We further excluded studies on prognostic factors on changes after memory training, as these were reviewed in another paper ([blinded for peer review]).

Prognostic model studies on healthy older participants (age ≥ 55 years) were included. Data from participants with mild cognitive impairment or dementia diagnosis, neurological and/or psychiatric diseases were excluded.

All prognostic models which investigate changes in memory test performance after memory training were included in the review. Memory training was defined as a CT that targets primarily on memory performance with a minimum of two sessions in total. The memory training can include paper-pencil or computerized tasks with clear cognitive rationale, which are administered either on personal devices or in individual- or group settings held by a facilitator. When cognitive multi-domain trainings were conducted, memory had to be the main component of the program (at least 50% of the exercises).

The included model studies had to investigate changes in verbal or non-verbal short- or long-term memory after memory training as an outcome, irrespectively whether it was assessed directly after the training and/or at FU. The outcomes had to be measured with established objective neuropsychological tests. We excluded subjective self-rated memory scales, as well as measures of memory strategy use. The factor measurement of the included studies had to be conducted before the memory training started, and there was no limitation regarding follow-up testing of outcomes.

The present review focuses on prognostic models for changes in memory performances after memory training only, due to different reasons: first, memory belongs to the most vulnerable cognitive functions in aging (e.g,. Salthouse, 2013). Second, as research is very limited so far in this field, we wanted to start with a rather narrow focus on a relevant field within the topic.

Data Extraction

Two review authors ([blinded for peer review]) independently extracted the data according to the Critical appraisal and data extraction for systematic reviews of prediction modelling studies (CHARMS) checklist (Moons et al., 2014) to investigate the quality of reporting of prognostic models.

Quality Assessment

Two reviewers ([blinded for peer review]) independently assessed the extracted studies for the risk of bias using the "Prediction model Risk of Bias Assessment Tool (PROBAST)" (Wolff et al., 2019) to examine the risk of bias in prognostic model studies across four domains: Participants, Predictors, Outcome, Analysis. Each of the domains was judged with "yes", "probably yes", "no", "probably no", and "no information". The studies were overall rated with low risk of bias, if all domains were rated low risk of bias. It was rated high risk of bias, if at least one domain was judged to be at high risk of bias or if a prediction model was developed without any external validation, and it was rated low risk of bias for all domains, it was downgraded to high risk of bias. A model without any external validation can only be considered as low risk of bias, if the development was based on a very large data set and included some form of internal validation (Wolff et al., 2019). Studies were rated as having an unclear risk of bias, if an unclear risk of bias was noted in at least one domain and it was low risk for all other domains.

Statistical analyses

In the pre-registration of the study, we registered a meta-analysis to investigate the predictive performance of the prognostic models. However, after the data extraction, we found that data on prognostic models of changes in memory test performance after memory training were too heterogeneous and based mostly on the same population (cf. 7 out of 12 studies reporting results of the ACTIVE trial) to conduct a meta-analysis.

RESULTS

Study selection

The total number of retrieved references and the numbers of included and excluded studies are documented in Figure 1 in a flow chart as recommended in the PRISMA statement (Moher et al., 2009). N = 10,703 studies were identified through the database search until October 2018 and by scanning the included studies in previously published systematic reviews and meta-analysis on memory training success in healthy older adults. N = 2,271 studies were identified in an update search in November 2019. After removing the duplicates, *n* = 9,979 studies were screened. We assessed 845 full-texts for eligibility. Finally, *n* = 12 studies were included in the present review. All studies were published in English.

Study characteristics

Table 1 gives an overview of the main characteristics of the included studies. Notably, *n* = 7 of the included studies investigated the same population (Gross et al., 2013; Gross & Rebok, 2011; Jones et al., 2013; Langbaum et al., 2009; Meyer et al., 2017; Rebok et al., 2013; Zahodne et al., 2015), namely the cognitive training trial ACTIVE.

The sample sizes varied between studies, ranging from *n* = 29 (Lovden, Brehmer, Li, & Lindenberger, 2012) to *n* = 703 (Gross et al., 2013; Gross & Rebok, 2011).
The mean age of the sample ranged from 66.90 years (Lovden et al., 2012) to 76.13 years (Macdonald, Stigsdotter-Neely, Derwinger, & Backman, 2006), with one study giving no data on the age of the memory training group (Zahodne et al., 2015). In most studies, the sample consisted of more female than male participants (overall: 71 % female). The samples were highly educated throughout the studies, ranging from a mean of 11.96 years of education (Macdonald et al., 2006) to a mean of 15.70 years (Zelinski, Peters, Hindin, Petway, & Kennison, 2014). The mean score of the cognitive screening instrument Mini Mental State

Examination (MMSE), which was assessed in seven studies at baseline to describe the baseline overall cognitive status of the study participants, has a maximum of 30 points indicating absolute cognitive health. The mean MMSE values of the study participants ranged from 27.00 points (Jones et al., 2013) to 28.90 points (McKitrick et al., 1999). All studies varied in their integration of different follow-up measurements with the $n = 7$ ACTIVE studies including most follow-up measurements: at one, two, three, five and ten years after intervention conduct, and $n = 3$ studies not assessing a follow-up measurement, but only a post-test measurement directly after the intervention (Beck et al., 2013; Lövdén et al., 2012, McKitrick et al, 1999).

A description of the different memory training interventions used (regarding main content, length, and frequency) is provided in Table 1.

Risk of Bias

Figure 2 displays the risk of bias rating of the included studies, assessed with the PROBAST tool (Wolff et al., 2019). Overall, the studies demonstrated a high risk of bias mainly due to the fact that their analysis was not conducted and/or reported according to the established guidelines and that internal and external model validation was missing. Only in the domain "Participants", all studies showed a low risk of bias rating.

Prognostic models of changes after memory training

Table 2 summarizes the analysis of methods and results of the included studies. Concerning statistical methods which are used in the included studies, six studies used a latent growth curve model to calculate their prognostic models (Gross et al., 2013; Gross & Rebok, 2011; Jones et al., 2013; Lovden et al., 2012; Rebok et al., 2013; Zahodne et al., 2015), four studies used a regression approach (Beck et al., 2013; Langbaum et al., 2009; McKitrick et al.,

1999; Meyer et al., 2017), one study used a multilevel modeling approach (Macdonald et al., 2006), and one study used structural equation modelling (Zelinski et al., 2014).

Over all models, the following predictors were investigated: Age (integrated in *n* = 11 prognostic models), sex (*n* = 8), education (*n* = 7), ethnicity (*n* = 6), neuropsychological baseline values at the beginning of the training (*n* = 6), self-rated health status (*n* = 4), depressive status (*n* = 1), socio-economic variables [i.e., living in major cities, neighborhood variables, employment status (*n* = 2)], and training related variables [length of training, type of pre-training (*n* = 1)].

The studies investigated verbal short- and long-term memory as well as non-verbal short-and long-term memory as primary outcomes. However, due to the fact that composite scores were build (*n* = 4 studies) or outcome parameters were not adequately described, a clear classification of outcome variables was difficult.

The numbers of predictors integrated in the prognostic models ranged from *n* = 1 (Jones et al., 2013, one predictor at several timepoints) to *n* = 15 (McKitrick et al., 1999). The predictors integrated in the model were highly heterogeneous: eight of twelve studies, however, integrated the sociodemographic predictors age, sex, and education in their models (with sometimes further additional predictors) (Beck et al., 2013; Gross et al., 2013; Gross & Rebok, 2011; Langbaum et al., 2009; Meyer et al., 2017; Rebok et al., 2013; Zahodne et al., 2015; Zelinski et al., 2014). In four of these studies (Meyer et al., 2017; Rebok et al., 2013; Zahodne et al., 2015; Zelinski et al., 2014), lower age and higher education predicted improvements in the memory outcomes (verbal short- and long-term memory) after training. However, it should be noted that three of these four studies are subsamples of the same study population of the ACTIVE trial (Meyer et al., 2017; Rebok et al., 2013; Zahodne et al., 2015). Female sex predicted gains in the memory outcome (Composite scores of verbal and non-verbal memory, separated for short- and long-term memory) after memory training in two of the investigated studies (Beck et al., 2013; Zahodne et al., 2015), yet, both studies integrated

also several further different predictors in the model (age, sex, education, ethnicity, health, depression vs. age, sex, education, marital status, baseline values, employment status). Three prognostic models found none of the investigated predictors (age, sex, and education as predictors in all three models; neuropsychological baseline values in two of the studies) to have a significant influence on the outcome (Beck et al., 2013; Gross et al., 2013; Gross & Rebok, 2011), indicating that all participants improved regardless of their individual characteristics.

DISCUSSION

This is the first review investigating prognostic models for changes in memory after memory training in healthy older adults. Our main finding is that although memory training has frequently been investigated in healthy older adults, only twelve studies so far exist which have published prognostic models; and notably, most of them (n = 7) are based on the same population (ACTIVE trial). Furthermore, our review indicates that the investigated models are highly heterogeneous regarding the number and the type of the prognostic factors as well as the statistical models. Finally, one result that has been found in several studies is that lower age combined with higher education seems to predict improvements in verbal short- and long-term memory after memory training over time. Furthermore, different statistical methods were used throughout the studies for calculating prognostic models and the overall reporting can be rated as deficient.

Identified predictors of changes after memory training

Results showed that in four of the included studies (Meyer et al., 2017; Rebok et al., 2013; Zahodne et al., 2015; Zelinski et al., 2014), lower age and higher education predicted improvements in the memory outcomes (verbal short- and long-term) after training; three of these studies are subsamples of the same study population of the ACTIVE trial (Meyer et al., 2017; Rebok et al., 2013; Zahodne et al., 2015). This result is contrary to findings from our

recently conducted review on prognostic factors of changes in memory after memory training in healthy older adults (Roheger, Folkerts, Krohm, Skoetz, & Kalbe, 2020), which shows that when using the change scores as the dependent variable in prognostic factor calculations, older participants benefit most from memory training. This result was discussed in terms of the compensation account, indicating that older participants may have more room for cognitive improvement (Lovden et al., 2012), while those who are already functioning at optimal levels have less room for changes in memory training performance. In both systematic reviews, the present at hand on prognostic models and the one on prognostic factors for changes after memory training ([blinded for peer review]), different types of memory trainings were investigated using either strategy-based or task-based trainings, individual- or group settings, paper-pencil or computerized exercise. Yet, no clear systematic pattern related to the investigated results could be found. For a better interpretation and a deeper understanding of the mechanisms of memory training, and for the future set-up of more individualized memory training approaches, a clear conceptualization of different memory training types should be designed, in which future memory studies could be clustered to shed further light on the differences of the direction of the prognostic factors in the two reviews. As "education" might be a proxy variable for e.g. socioeconomic status, early life factors, occupational health, or even the willingness to engage in lifelong learning or new activities (Krieger, Williams, & Moss, 1997), integrating education in the prognostic model could have a further impact on all other investigated variables, maybe even explaining the observed differences in the "age" variable throughout studies (as in [blinded for peer review]). Different results may be due to the impact of other prognostic factors in the model, leading to a different weighting of the prognostic factors in the models compared to single prognostic factor studies. Therefore, it is of high importance to evaluate prognostic factors in a stepwise modulation process, and not integrate all possible prognostic factors at once in a model at hand, especially when no cross-validation can be done and it is not known whether and how the single prognostic factors explain variance

in the models. Further, it should be noted that interpreting results of studies that are subsamples of the same study population is always complex, as the samples are not independent. Instead of creating subsamples to investigate different models, subsamples should be used to cross-validate the found results in a similar prognostic model. Further, to ensure a high research quality, specific a-priori hypothesis about prognostic models results should be stated.

Two of the studies included in our review showed that female sex predicted gains in the memory outcome after memory training (Beck et al., 2013; Zahodne et al., 2015), fitting to the notion of sex-specific plasticity (Beinhoff et al.,2008). This result is also supported by a study of Munro et al. (2012) showing that healthy older female participants perform better on tests of memory and verbal learning than men in general (Munro et al., 2012). However, in this study no memory training was conducted. A study by Rahe et al. (2015) could show that after a CT female patients with mild cognitive impairment (MCI) showed stronger improvements after the training in the domains delayed verbal episodic memory, and working memory (Rahe et al., 2015). While further studies are needed to elucidate this topic in more detail, it could be possible that women´s larger gains delayed verbal episodic memory tasks after CT might be easier to find in patients with cognitive decline, including MCI and Alzheimer´s disease (Beinhoff et al., 2008). Yet, it is important to be aware that these sex differences often have small effect sizes and further research is urgently needed, especially in healthy older participants in the context of CT (Choleris, Galea, Sohrabji, & Frick, 2018).

Three models found none of the investigated predictor to have a significant impact on changes after memory training when including amongst others age, sex, education, neuropsychological baseline variables (Beck et al., 2013; Gross et al., 2013; Gross & Rebok, 2011), which indicates that training gains were independent of specific prognostic factors. Yet, two of these studies are again a sub-cohort of the ACTIVE trial (Gross et al., 2013; Gross & Rebok, 2011), which showed significant prognostic factors in other investigated models. Therefore, it is possible that results are obliterated by a specific sample selection.

Summarized, data is highly heterogeneous regarding investigated predictors in the prognostic models on the one hand, and on the other hand only of limited explanatory power, as seven of the studies are based on the same population (ACTIVE trial). We could not find a clear pattern with regard the memory training content. More studies are needed including robust a-priori hypotheses with a profound theoretical basis and internal and external model validation processes to strengthen results.

Identified statistical methods used for prognostic models

The representation and measurement of change is a fundamental concern in scientific disciplines, as longitudinal research designs pose several unique problems because they involve variables with correlated observations (Duncan & Duncan, 2004). Therefore, it is stated that an appropriate developmental model is one that not only describes a single individual's developmental trajectory, but that also integrated individual differences in these trajectories over a period of time (Duncan & Duncan, 2004). In the investigated studies, different statistical methods were used to calculate prognostic models for changes after memory training, namely structural equation models (especially latent growth curve models), regression models, and multilevel models.

Multiple regression models, as well as analyses of variance (which Cohen demonstrated in 1968 are essentially identical data analytic systems (Cohen, 1968)), mainly focus on differences in mean changes instead of intra-individual variability and growth trajectories (Voelkle, 2007). *Latent growth curve models*, on the other side (which belong to the family of structural equation models), are interpreted as individual differences in factors of growth trajectories over time (mainly the rates of changes and initial status), meaning that it allows for the study of individual differences in the parameters that control the pattern of growth over time – on the group and individual level (McArdle, 1988). Further, predictors of these differences can be studied to answer which variables explain effects on the rate of

development. Even though there was a long debate on which model is “more appropriate” to model change, Voelkle (2007) could show that both approaches are essentially identical, and that multiple regression models are special cases of the more general latent growth curve approach (Voelkle, 2007). *Multilevel models* (which are also known as hierarchical linear models, mixed models, or random-effect models) answers similar questions as the latent growth curve modelling approach (Raudenbush & Bryk, 2002) and are widely seen as an “improvement” compared with classical regression models as they give more accurate predictions than the no-pooling or complete-data-pooling regressions (Gelman, 2006).

Summarized, latent growth curve model and multilevel model approaches seem to be the most appropriate to model predictors of change over time, even though also multiple regression models can lead to similar results when meeting specific assumptions (e.g., the choice of an adequate dependent variable as the choice of the dependent variables [change scores vs. raw scores] may influence the direction of the results in multiple regression analyses but not in other statistical model approaches as they modulate their dependent variables in a different way; for a further discussion on dependent variables in multiple regression analyses see [blinded for peer review]). Therefore, all investigated studies in the systematic review used appropriate modeling approaches. Even though the overall reporting quality of the studies was quite high, future studies could be more precise in the correct and consistent naming of the modeling techniques they have used and provide detailed descriptions why they have chosen a specific modeling approach. Further, especially in complex modeling approaches, results should not solely be presented in statistical language, but filled with results with regard to content and examples in order to help the reader to better understand the specific results and interpretations of the prognostic models. Yet, all statistical models should be validated either by internal validation, external validation, or temporal validation (Altman et al., 2009).

Limitations of the present systematic review

Some limitations have to be taken into account when interpreting the results of the present review. First, it was difficult in the study search process to distinguish between factor finding and prognostic model studies, as the statistical methods were often not clearly reported so that in some cases it was not possible to determine which prognostic variables were used in the final calculations. Therefore, it might be possible that studies were not correctly classified and studies, which would have been within the scope of the review, were excluded or investigated in the review on prognostic factors due to incomprehensive statistical analyses resulting in only a few investigated studies in the present review.

Further, interpretation of the results was difficult as seven of the included studies were based on the same population (partly only subsamples were used) and a summary of the results may therefore be not representative or redundant. None of the included prognostic model studies conducted an external model validation and therefore results may be insufficient. In the present review, we only included studies in English or German language, so that we may therefore have missed studies published in other languages. The present systematic review only focuses on memory outcomes after memory training, hereby disregarding other cognitive domains, as well as other non-cognitive outcomes (e.g., depression, quality of life, activities of daily living). Further systematic reviews are needed to elaborate the knowledge on prognostic models of CT success. Yet, the present review can be seen as a starting signal for further and more accurate research and reporting on prognostic models studies for changes after memory training.

As a final limitation, we could not perform a meta-analysis on the investigated prognostic models as planned and stated in the pre-registration of the present systematic review (ID: [blinded for peer review]) due to the heterogeneity of the investigated models and the fact that most studies were based on the same population, which would have led to distorted results.

Strengths of the present systematic review

This is the first review dealing with prognostic models for changes after memory training in healthy older adults highlighting not only the statistical modeling approaches used, but also the need for further and theory-based prognostic model assumptions and validation of currently existing models. A further strength of the review is that it was conducted using Cochrane standards, and that the search was conducted in several databases to ensure an exhausting overview of this important research topic.

Implications and Conclusion

Only a few studies investigate prognostic models of changes after memory training, most of which are based on the same study population so that no clear pattern could be detected. Overall, the investigated model studies showed high risks of bias ratings and a clear need for a better reporting of their used statistical methods and the need for internal and external model validation. Therefore, more prognostic model studies are needed, which are not only well reported in their design, but also cross-validated to ensure a high research quality. As prognostic model studies are of high importance regarding an individual prevention approach of cognitive decline in higher age, further research is urgently needed.

Table 1. Participants' Demographics and Memory Training Characteristics

Study	**Study name**	**Participants**					**Training**			**Psychoeducation**	**Digital**	**Group Setting**	**Strategy training**	**Specific prognostic trial**
Author (year)		**n**[a]	**Age** (in years)	**Sex**	**Education** (in years)	**Global Cognition** (at baseline)	**Total Time** (in minutes)	**Frequency**	**Description of Training**					
ACTIVE studies														
Gross & Rebok, 2011	ACTIVE	703	73.53 (6.02)	♂ = 24% ♀ = 76%	13.59 (2.73)	MMSE: 27.29 (2.05)	750	10 weekly training sessions.	Teaching and practicing several mnemonic strategies (Method of Loci, Association, Visualization).	X		X	X	
Gross et al., 2013	ACTIVE	703	73.53 (6.02)	♂ = 24% ♀ = 76%	13.59 (2.73)	MMSE: 27.29 (2.05)	750	10 weekly training sessions.	Teaching and practicing several mnemonic strategies (Method of Loci, Association, Visualization).	X		X	X	
Jones et al., 2013	ACTIVE	296	74.00 (6.00)	♂ = 21% ♀ = 79%	13.00 (3.00)	MMSE: 27.00 (2.00)	750	10 weekly training sessions.	Teaching and practicing several mnemonic strategies (Method of Loci, Association, Visualization).	X		X	X	
Langbaum et al., 2009	ACTIVE	619	73.40 (5.90)	♂ = 23% ♀ = 77%	13.60 (2.70)	MMSE: 27.40 (2.00)	750	10 weekly training sessions.	Teaching and practicing several mnemonic strategies (Method of Loci, Association, Visualization).	X		X	X	
Meyer et al., 2017	ACTIVE	624	73.50 (6.00)	♂ = 24% ♀ = 76%	13.53 (2.69)	/	750	10 weekly training sessions.	Teaching and practicing several mnemonic strategies (Method of Loci, Association, Visualization).	X		X	X	
Rebok et al., 2013	ACTIVE	629	73.50 (6.00)	♂ = 23% ♀ = 77%	13.70 (2.70)	MMSE: 27.30 (2.00)	750	10 weekly training sessions.	Teaching and practicing several mnemonic strategies (Method of Loci, Association, Visualization).	X		X	X	
Zahodne et al., 2015	ACTIVE	693	/	/	/	/	750	10 weekly training sessions.	Teaching and practicing several mnemonic strategies (Method of Loci, Association, Visualization).	X		X	X	
Studies based on different populations														
Beck et al., 2013	SeniorWISE	116	71.90 (6.60)	♂ = 23% ♀ = 77%	/	MMSE: 28.40 (1.50)	720	12 weekly 1h sessions.	Psychoeducation and training on different memory strategies and problem solving.	X		X	X	
Lovden et al., 2012	/	29	66.90 (3.70)	♂ = 52% ♀ = 48%	/	/	/	3 – 7 training sessions.	Maximum of 36 lists of adaptive practice of location-word pairs.		X		X	X
Macdonald et al., 2006	/	97	Young-old: 64.43 (2.47) Old-old: 76.13 (4.09)	Young-old: ♂ = 47% ♀ = 53% Old-old: ♂ = 33% ♀ = 67%	Young-old: 11.96 (3.34) Old-old: 12.19 (3.85)	/	600 + 240	10 sessions twice a week. + 4 introductory sessions in group settings.	Number-Consonant mnemonic strategies were taught in sessions 1 – 4. Memorization of list of four-digit numbers in sessions 6 – 16.		X	X	X	X
McKitrick et al., 1999	Population of Brooks, Friedman, Pearman, Gray, & Yesavage, 1999	224	68.60 (7.00)	♂ = 30% ♀ = 70%	15.31 (2.51)	MMSE: 28.90 (n.a.)	/	5 daily 2h sessions over 2 weeks.	Two mnemonic techniques were taught (name-face and Method of Loci).	?	?	?	X	?
Zelinski et al., 2014	IMPACT	242	75.60 (6.60)	♂ = 42% ♀ = 58%	15.70 (2.60)	/	2400	5 daily 1h sessions.	Computerized cognitive training program on the speeded auditory discrimination task.		X			X

Note. [a] = only participants in the memory training group. Abbreviations: MMSE = Mini Mental State Examination.

Table 2. Prognostic Analysis: Analyses, Outcomes, Results, and Timing

Study	Analysis[a]	Prognostic Variables	Outcome(s)	Prediction Results				
			Changes in…	Age	Sex	Education	Baseline performance	Others
ACTIVE studies								
Gross & Rebok, 2011	Parallel process latent growth curve models	Age Sex Education Self-rated Health Status Ethnicity	HVLT Clustering Scores	X	X	X	--	X
			AVLT Clustering Scores	X	X	X	--	X
Gross et al., 2013	Latent growth curve model	Age Sex Education Self-rated Health Status Ethnicity Baseline Clustering	HVLT Learning Curve	X	X	X	X	X
			AVLT Learning Curve	X	X	X	X	X
Jones et al., 2013	Random effects growth curve model	Age	Memory composite score for verbal and non-verbal measures (HVLT, AVLT, RBPRT)	↑ (slower decline)	--	--	--	--
Langbaum et al., 2009	Univariate and multiple polytomous logistic regressions	Age Sex Education Ethnicity Global status at baseline Memory Baseline performance	HVLT respondents (conversion rate from people who improve or do not improve in the HVLT)	X	X	↑	X	Memory baseline performance ↑ Reasoning baseline performance ↑ Speed of processing ability ↓ Ethnicity ↑
			AVLT respondents (conversion rate from people who improve or do not improve in the HVLT)	X	X	↑	↓	X
Meyer et al., 2017	Mixed effect regression models	Time Ethnicity Age Education Intervention Initial Gain Neighborhood socioeconomic position Major City	Memory composite score for verbal and non-verbal measures (HVLT, AVLT, RBPRT)	↓*	--	↑*	--	Time ↓* Ethnicity (Black Race) ↓* Initial Gain ↑*

		Percentage Minority						
Rebok et al., 2013	Multiple Group Latent Growth Curve Models	Age Sex Ethnicity Self-rated Health Status Education	Memory composite score for verbal and non-verbal measures (HVLT, AVLT, RBPRT)	↓	X	↑	X	Self-rated health ↑
Zahodne et al., 2015	Latent Growth Curve Models	Age Sex Ethnicity Education Self-rated Health Status Depression	Memory composite score for verbal and non-verbal measures (HVLT, AVLT, RBPRT)	↓*	↑* (female sex)	↑*	--	Ethnicity ↓* Health ↓ Depression ↓*
Studies based on different populations								
Beck et al., 2013	General linear mixed model approach	Baseline values Age Sex Marital status Education Employment status	RBANS indices: Immediate Recall	X	↑* (female sex)	X	X	X
			RBANS indices: Delayed Recall	X	X	X	X	X
Lovden et al., 2012	Confirmatory two-factor model latent curve model	Baseline performance	Timed recalled score: instruction gains	--	--	--	↓	--
		Baseline performance	Practice Gains in immediate word list recall	--	--	--	X	--
Macdonald et al., 2006	Multilevel modelling	1. Young-old: Time Perceptual speed Episodic memory Working memory	Forgetting in immediate word list recall	--	--	--	--	Time ↓* Perceptual speed ↓ Episodic memory ↑* Working memory ↓
		2. Old-old: Time Perceptual speed Episodic memory Working memory	Forgetting in immediate word list recall	--	--	--	--	Time ↓* Perceptual speed ↑ Episodic memory ↓ Working memory ↑*

McKitrick et al., 1999	Logistic regression	PASAT SDM TMT OR QV MMSE LG AL RBVRT RCPM Age Sex Length of training Type of pre-training Baseline performance	Word recall (immediate recall)	X	X	--	↑*	OR↓* AL↓*
			Name recall (immediate recall)	X	X	--	↑*	AL↓*
Zelinski et al., 2014	Structural Equation Modeling	Age Sex Education	Time order judgement sound sweep discrimination task	X	X	↓	--	--
			Recognition of syllables	↓*	X	↑*	--	--

Note. Abbreviations: RBANS = Repeatable Battery for the Assessment of Neuropsychological Status. HVLT = Hopkins Verbal Learning Test. AVLT = Auditory Verbal Learning Test. RBPRT = Rivermead Behavioral Paragraph Recall Test. PASAT = Paced Auditory Serial Addition Test. SDM = Symbol-Digit Modalities. TMT = Trail Making Test. OR = Object Rotation. QV = Quick Vocabulary. MMSE = Mini Mental State Evaluation. LM = Logical Memory. AL = Associate Learning. X = predictor was investigated, but had no significant effect. -- = predictor was not investigated in the study, ↑ = higher predictor scores show improvement in the outcome domain. ↓ = lower predictor scores show improvement in the outcome domain. * = Results are reported as significant. [a] Terms are used as described in the studies.

Figure 1: PRISMA flow diagram

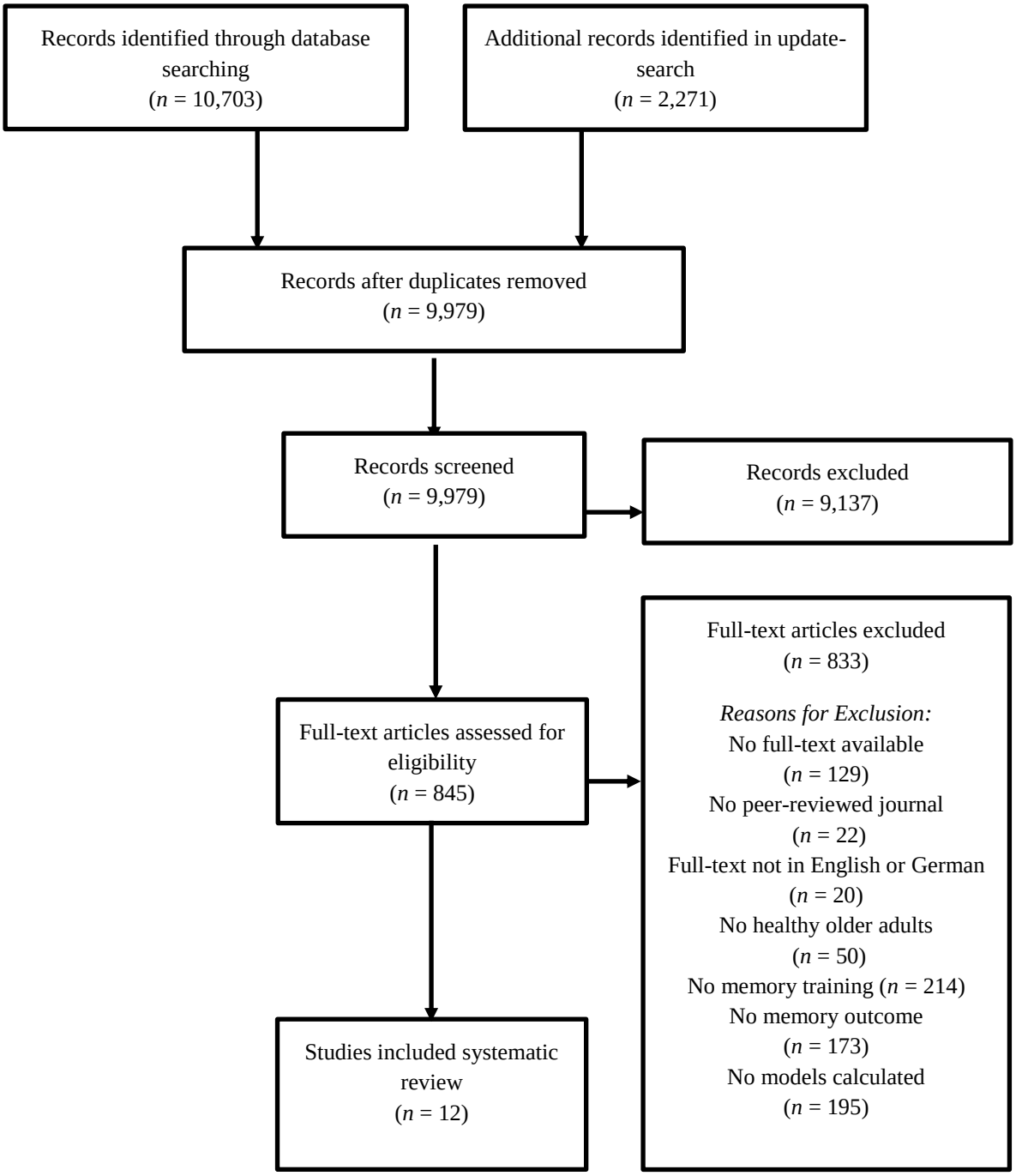
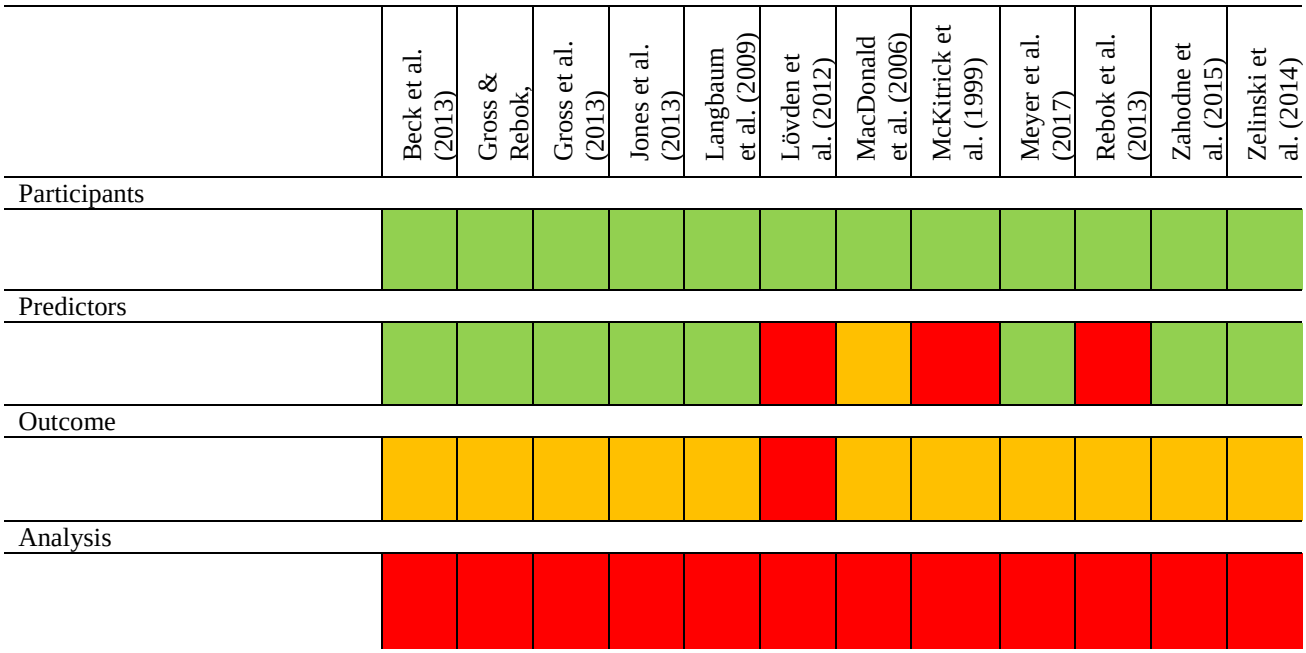


Figure 2: Risk of Bias



Note. Risk of bias assessment using the "Prediction model Risk of Bias Assessment Tool (PROBAST)" (Wolff et al., 2019) to examine the risk of bias in prognostic factors studies across four domains: Participants, Predictors, Outcome, Analysis. Each of the domains was judged with "low risk" (depicted in green), "high risk"(red), "unclear risk of bias" (yellow).

References

Altman, D. G., Vergouwe, Y., Royston, P., & Moons, K. G. M. (2009). Prognosis and prognostic research: Validating a prognostic model. *BMJ (Clinical Research Ed.)*, *338*, b605. https://doi.org/10.1136/bmj.b605

Beck, C., Fausett, J. K., Krukowski, R. A., Cornell, C. E., Prewitt, T. E., Lensing, S., . . . West, D. S. (2013). A randomized trial of a community-based cognitive intervention for obese senior adults. *Journal of Aging and Health*, *25*(1), 97–118. https://doi.org/10.1177/0898264312467374

Brooks, J. O., Friedman, L., Pearman, A. M., Gray, C., & Yesavage, J. A. (1999). Mnemonic training in older adults: Effects of age, length of training, and type of cognitive pretraining. *International Psychogeriatrics*, *11*(1), 75–84. https://doi.org/10.1017/s1041610299005608

Chiu, H.-L., Chu, H., Tsai, J.-C., Liu, D., Chen, Y.-R., Yang, H.-L., & Chou, K.-R. (2017). The effect of cognitive-based training for the healthy older people: A meta-analysis of randomized controlled trials. *PloS One*, *12*(5), e0176742. https://doi.org/10.1371/journal.pone.0176742

Choleris, E., Galea, L. A. M., Sohrabji, F., & Frick, K. M. (2018). Sex differences in the brain: Implications for behavioral and biomedical research. *Neuroscience and Biobehavioral Reviews*, *85*, 126–145. https://doi.org/10.1016/j.neubiorev.2017.07.005

Cohen, J. (1968). Multiple regression as a general data-analytic system. *Psychological Bulletin*, *70*(6, Pt.1), 426–443. https://doi.org/10.1037/h0026714

Duncan, T. E., & Duncan, S. C. (2004). An introduction to latent growth curve modeling. *Behavior Therapy*, *35*(2), 333–363. https://doi.org/10.1016/s0005-7894(04)80042-x

Gelman, A. (2006). Multilevel (Hierarchical) Modeling: What It Can and Cannot Do. *Technometrics.* Advance online publication. https://doi.org/10.1198/004017005000000661

Gross, A. L., & Rebok, G. W. (2011). Memory training and strategy use in older adults: Results from the ACTIVE study. *Psychology and Aging*, *26*(3), 503–517. https://doi.org/10.1037/a0022687

Gross, A. L., Rebok, G. W., Brandt, J., Tommet, D., Marsiske, M., & Jones, R. N. (2013). Modeling learning and memory using verbal learning tests: Results from ACTIVE. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *68*(2), 153–167. https://doi.org/10.1093/geronb/gbs053

Hitchcock, C., Werner-Seidler, A., Blackwell, S. E., & Dalgleish, T. (2017). Autobiographical episodic memory-based training for the treatment of mood, anxiety and stress-related disorders: A systematic review and meta-analysis. *Clinical Psychology Review*, *52*, 92–107. https://doi.org/10.1016/j.cpr.2016.12.003

Jockwitz, C., Mérillat, S., Liem, F., Oschwald, J., Amunts, K., Caspers, S., & Jäncke, L. (2019). Generalizing age effects on brain structure and cognition: A two-study comparison approach. *Human Brain Mapping*, *40*(8), 2305–2319. https://doi.org/10.1002/hbm.24524

Jones, R. N., Marsiske, M., Ball, K., Rebok, G., Willis, S. L., Morris, J. N., & Tennstedt, S. L. (2013). The ACTIVE cognitive training interventions and trajectories of performance among older adults. *Journal of Aging and Health*, *25*(8 Suppl), 186S-208S. https://doi.org/10.1177/0898264312461938

Krieger, N., Williams, D. R., & Moss, N. E. (1997). Measuring social class in US public health research: Concepts, methodologies, and guidelines. *Annual Review of Public Health, 18*, 341–378. https://doi.org/10.1146/annurev.publhealth.18.1.341

Langbaum, J. B. S., Rebok, G. W., Bandeen-Roche, K., & Carlson, M. C. (2009). Predicting memory training response patterns: Results from ACTIVE. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *64*(1), 14–23. https://doi.org/10.1093/geronb/gbn026

Lovden, M., Brehmer, Y., Li, S.-C., & Lindenberger, U. (2012). Training-induced compensation versus magnification of individual differences in memory performance. *Frontiers in Human Neuroscience*, *6*, 141. https://doi.org/10.3389/fnhum.2012.00141

Macdonald, S. W. S., Stigsdotter-Neely, A., Derwinger, A., & Backman, L. (2006). Rate of acquisition, adult age, and basic cognitive abilities predict forgetting: New views on a classic problem. *Journal of Experimental Psychology. General*, *135*(3), 368–390. https://doi.org/10.1037/0096-3445.135.3.368

McArdle, J. (1988). Dynamic but Structural Equation Modeling of Repeated Measures Data. In *Handbook of Multivariate Experimental Psychology* (pp. 561–614). Springer, Boston, MA. https://doi.org/10.1007/978-1-4613-0893-5_17

McKitrick, L. A., Friedman, L. F., Brooks, J. O., Pearman, A., Kraemer, H. C., & Yesavage, J. A. (1999). Predicting response of older adults to mnemonic training: Who will benefit? *International Psychogeriatrics*, *11*(3), 289–300. https://doi.org/10.1017/s1041610299005852

Meyer, O. L., Sisco, S. M., Harvey, D., Zahodne, L. B., Glymour, M. M., Manly, J. J., & Marsiske, M. (2017). Neighborhood Predictors of Cognitive Training Outcomes and Trajectories in ACTIVE. *Research on Aging*, *39*(3), 443–467. https://doi.org/10.1177/0164027515618242

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Journal of Clinical Epidemiology*, *62*(10), 1006–1012. https://doi.org/10.1016/j.jclinepi.2009.06.005

Moons, K. G. M., Groot, J. A. H. de, Bouwmeester, W., Vergouwe, Y., Mallett, S. [Susan], Altman, D. G., . . . Collins, G. S. (2014). Critical appraisal and data extraction for systematic reviews of prediction modelling studies: The CHARMS checklist. *PLoS Medicine*, *11*(10), e1001744. https://doi.org/10.1371/journal.pmed.1001744

Munro, C. A., Winicki, J. M., Schretlen, D. J., Gower, E. W., Turano, K. A., Muñoz, B., . . . West, S. K. (2012). Sex Differences in Cognition in Healthy Elderly Individuals. *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition*, *19*(6), 759–768. https://doi.org/10.1080/13825585.2012.690366

Rahe, J., Liesk, J., Rosen, J. B., Petrelli, A., Kaesberg, S., Onur, O. A., . . . Kalbe, E. [Elke] (2015). Sex differences in cognitive training effects of patients with amnestic mild cognitive impairment. *Neuropsychology, Development, and Cognition. Section B, Aging,*

Neuropsychology and Cognition, *22*(5), 620–638. https://doi.org/10.1080/13825585.2015.1028883

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*: SAGE.

Rebok, G. W., Langbaum, J. B. S., Jones, R. N., Gross, A. L., Parisi, J. M., Spira, A. P., . . . Brandt, J. (2013). Memory training in the ACTIVE study: How much is needed and who benefits? *Journal of Aging and Health*, *25*(8 Suppl), 21S-42S. https://doi.org/10.1177/0898264312461937

Roheger, M., Folkerts, A. K., Krohm, F., Skoetz, N., & Kalbe, E. [E.] (2020). Prognostic factors for change in memory test performance after memory training in healthy older adults: A systematic review and outline of statistical challenges. *Diagnostic and Prognostic Research*.

Roozenbeek, B., Maas, A. I. R., Lingsma, H. F., Butcher, I., Lu, J., Marmarou, A., . . . Steyerberg, E. W. (2009). Baseline characteristics and statistical power in randomized controlled trials: Selection, prognostic targeting, or covariate adjustment? *Critical Care Medicine*, *37*(10), 2683–2690. https://doi.org/10.1097/ccm.0b013e3181ab85ec

Rosi, A., Del Signore, F., Canelli, E., Allegri, N., Bottiroli, S., Vecchi, T., & Cavallini, E. (2018). The effect of strategic memory training in older adults: Who benefits most? *International Psychogeriatrics*, *30*(8), 1235–1242. https://doi.org/10.1017/S1041610217002691

Salthouse, T. A. (2006). Mental exercise and mental aging: Evaluating the validity of the "use it or lose it" hypothesis. *Perspectives on Psychological Science*, *1*(1), 68-87.

Simon, S. S., Tusch, E. S., Feng, N. C., Håkansson, K., Mohammed, A. H., & Daffner, K. R. (2018). Is Computerized Working Memory Training Effective in Healthy Older Adults? Evidence from a Multi-Site, Randomized Controlled Trial. *Journal of Alzheimer's Disease : JAD*, *65*(3), 931–949. https://doi.org/10.3233/JAD-180455

Steyerberg, E. W., Moons, K. G. M., van der Windt, D. A., Hayden, J. A., Perel, P., Schroter, S., . . . Altman, D. G. (2013). Prognosis Research Strategy (PROGRESS) 3: Prognostic model research. *PLoS Medicine*, *10*(2), e1001381. https://doi.org/10.1371/journal.pmed.1001381

Verhaeghen, P., Geraerts, N., & Marcoen, A. (2000). Memory complaints, coping, and well-being in old age: A systemic approach. *The Gerontologist*, *40*(5), 540–548. https://doi.org/10.1093/geront/40.5.540

Voelkle, M. C. (2007). Latent growth curve modeling as an integrative approach to the analysis of change. *Psychology Science*, p. 375-414. Retrieved from https://scholar.google.com/scholar?hl=de&as_sdt=0%2C5&q=Latent+growth+curve+modeling+as+an+integrative+approach+to+the+analysis+of+change&btnG=

Wolff, R. F., Moons, K. G. M., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., . . . Mallett, S. [Sue] (2019). Probast: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Annals of Internal Medicine*, *170*(1), 51–58. https://doi.org/10.7326/M18-1376

Zahodne, L. B., Meyer, O. L., Choi, E., Thomas, M. L., Willis, S. L., Marsiske, M., . . . Parisi, J. M. (2015). External locus of control contributes to racial disparities in memory

and reasoning training gains in ACTIVE. *Psychology and Aging*, *30*(3), 561–572. https://doi.org/10.1037/pag0000042

Zelinski, E. M., Peters, K. D., Hindin, S., Petway, K. T., & Kennison, R. F. (2014). Evaluating the relationship between change in performance on training tasks and on untrained outcomes. *Frontiers in Human Neuroscience*, *8*, 617. https://doi.org/10.3389/fnhum.2014.00617