Prognostic Models for Changes in Memory Performance After Memory Training in Healthy Older Adults: a Systematic Review

Mandy Roheger[1,2] · Ann-Kristin Folkerts[1] · Fabian Krohm[1] · Nicole Skoetz[3] · Elke Kalbe[1]

Abstract

Identifying individuals' profiles of prognostic factors that predict improvements after nonpharmacological interventions such as memory trainings may help to not only predict individuals' future outcomes after such intervention, but also 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. 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. 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 higher improvements after memory training. 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

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;

✉ Elke Kalbe
elke.kalbe@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

2   Department of Neurology, University Medicine Greifswald, Greifswald, Germany

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

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 predict individuals' future outcome after CT. Further, it may improve informed decision-making among clinicians to follow a personalized medicine approach (Altman et al. 2009). It can also be used to improve the design and analysis of randomized 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 et al. 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 some studies showed that it can maintain further non-trained memory functions as well as non-cognitive abilities in older adults (Hitchcock et al. 2017; Rosi et al., 2018; Simon et al., 2018). However, transfer of memory training to non-trained functions is limited. Notably, results from the literature indicates that there is a great variability of responsiveness among healthy older training participants, e.g., with some studies showing that participants with older age benefit most from training (Brooks et al., 1999), whereas other studies show that younger participants benefit most from training (Langbaum et al., 2009). 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 et al., 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. Furthermore, prognostic models provide different information than prognostic factor studies, and have to be assessed with different tools regarding risk of bias judgment. 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 et al., 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 (CRD42018105803, https://www.crd.york.ac.uk/PROSPERO).

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 2-week time frame. The full search strings for each database are presented in the Supplementary Material, Tables 3–6.

Two review authors (MR, AKF) 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 (EK) 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 (Roheger et al., 2020).

Prognostic model studies on healthy older participants (age ≥ 55 years) were included. Data from participants with mild cognitive impairment or dementia diagnosis and 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 (MR, AKF) independently extracted the data according to the critical appraisal and data extraction for systematic reviews of prediction modeling studies (CHARMS) checklist (Moons et al., 2014) to investigate the quality of reporting of prognostic models.

Quality Assessment

Two reviewers (MR, AKF) 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 all other domains were rated as low risk of bias. A model without any external validation can only be considered 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 Fig. 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. $N = 2271$ studies were identified in an update search in November 2019. After removing the duplicates, $n = 9979$ 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 $n = 29$ (Lövdén et al., 2012) to $n = 703$ (Gross et al., 2013; Gross & Rebok, 2011).

The mean age of the sample ranged from 66.90 years (Lövdén et al., 2012) to 76.13 years (Macdonald et al., 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 et al., 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 1, 2, 3, 5, and 10 years after intervention conduct, and $n = 3$ studies not assessing a follow-up measurement, but only a post-test
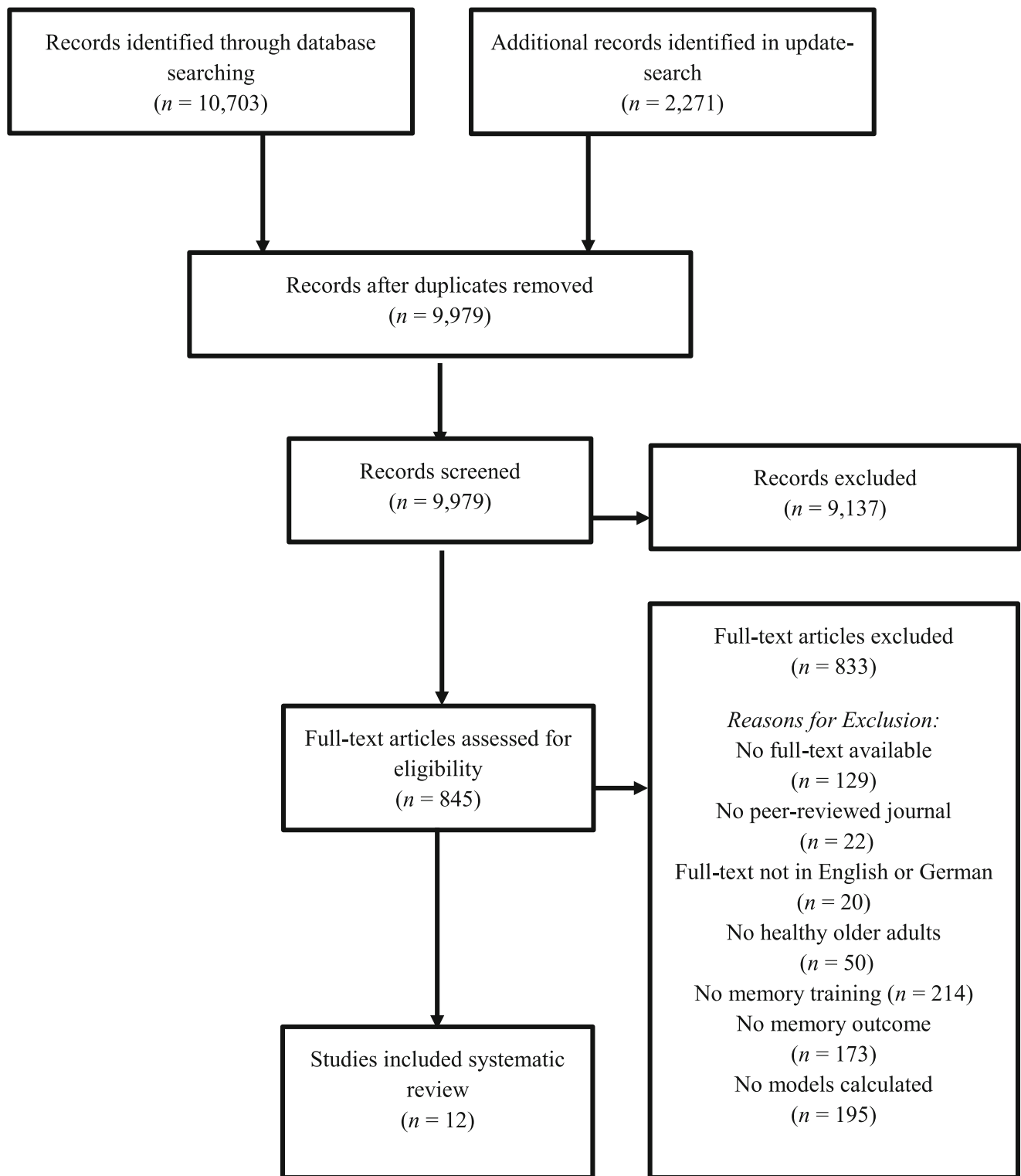
```
┌─────────────────────────────┐   ┌─────────────────────────────┐
│ Records identified through  │   │ Additional records          │
│ database searching          │   │ identified in update-       │
│ (n = 10,703)                │   │ search                      │
│                             │   │ (n = 2,271)                 │
└─────────────────────────────┘   └─────────────────────────────┘
```

Records after duplicates removed
(*n* = 9,979)

Records screened
(*n* = 9,979)

Records excluded
(*n* = 9,137)

Full-text articles assessed for
eligibility
(*n* = 845)

Full-text articles excluded
(*n* = 833)

Reasons for Exclusion:
No full-text available
(*n* = 129)
No peer-reviewed journal
(*n* = 22)
Full-text not in English or German
(*n* = 20)
No healthy older adults
(*n* = 50)
No memory training (*n* = 214)
No memory outcome
(*n* = 173)
No models calculated
(*n* = 195)

Studies included systematic
review
(*n* = 12)

Fig. 1 PRISMA flow diagram

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

Table 1 Participants' demographics and memory training characteristics

Author (year)	Study name	Participants n[a]	Age (in years)	Sex	Education (in years)	Global cognition (at baseline)	Training Total time (in minutes)	Frequency	Description of training	Psychoeducation	Digital	Group setting	Strategy training	Specific prognostic trial
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,	X		X	X	

Table 1 (continued)

Study name	Author (year)	Participants n[a]	Age (in years)	Sex	Education (in years)	Global cognition (at baseline)	Training Total time (in minutes)	Frequency	Description of training	Psychoeducation	Digital	Group setting	Strategy training	Specific prognostic trial
									association, visualization)					
ACTIVE	Zahodne et al., 2015	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														
SeniorWISE	Beck et al., 2013	116	71.90 (6.60)	♂ = 23% ♀ = 77%	/	MMSE: 28.40 (1.50)	720	12 weekly 1 h sessions	Psychoeducation and training on different memory strategies and problem-solving	X		X	X	
/	Lövdén 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
Population of Brooks, Friedman, Pearman, Gray, & Yesavage, 1999	McKitrick et al., 1999	224	68.60 (7.00)	♂ = 30% ♀ = 70%	15.31 (2.51)	MMSE: 28.90 (n.a.)	/	5 daily 2 h sessions over 2 weeks	Two mnemonic techniques were taught (name-face and method of loci)	?	?	?	X	?
IMPACT	Zelinski et al., 2014	242	75.60 (6.60)	♂ = 42% ♀ = 58%	15.70 (2.60)	/	2400	5 daily 1 h sessions	Computerized cognitive training program on the speeded auditory		X			X

Table 1 (continued)

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					
									discrimination task					

[a] Only participants in the memory training group. *MMSE* Mini Mental State Examination

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; Lövdén 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 modeling (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 ($n = 1$), socioeconomic 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 built ($n = 4$ 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

Fig. 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)

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 et al., 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 (Lövdén 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 (Roheger et al., 2020), different types of memory trainings were investigated using either strategy-based or task-based trainings, individual or group settings, or paper-pencil or computerized exercise.

Table 2 Prognostic analysis: analyses, outcomes, results, and timing

Study	Analysis[a]	Prognostic variables	Outcome(s) Changes in…	Prediction results				
				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 AVLT clustering scores	X X	X X	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 AVLT learning curve	X X	X X	X 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) AVLT respondents (conversion rate from people who improve or do not improve in the HVLT)	X X	X X	↑ ↑	X →	Memory baseline performance ↑ Reasoning baseline performance ↑ Speed of processing ability ↑ Ethnicity ↑ X
Meyer et al., 2017	Mixed effect regression models	Time Ethnicity Age Education Intervention Initial Gain Neighborhood socioeconomic position Major City Percentage Minority	Memory composite score for verbal and non-verbal measures (HVLT, AVLT, RBPRT)	↓*	--	↑*	--	Time ↓* Ethnicity (Black race) ↓* Initial gain ↑*
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	Memory composite score for verbal and non-verbal measures (HVLT, AVLT, RBPRT)	↓*	↑* (female sex)	↑*	--	Ethnicity ↓* Health ↓ Depression ↓*

Table 2 (continued)

Study	Analysis[a]	Prognostic variables	Outcome(s) Changes in…	Prediction results				
				Age	Sex	Education	Baseline performance	Others
		Education Self-rated health status Depression	measures (HVLT, AVLT, RBPRT)					
Studies based on different populations								
Beck et al., 2013	General linear mixed model approach	Baseline values Age Sex Marital status	RBANS indices: immediate recall	X	↑ * (female sex)	X	X	X
		Education Employment status	RBANS indices: delayed recall	X	X	X	X	X
Lövdén 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 modeling	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 SDMT TMT OR QV MMSE LGAL RBVRT RCPM	Word recall (immediate recall)	X	X	--	↑ *	OR↓* AL↓*
		Age Sex Length of training Type of pre-training Baseline performance	Name recall (immediate recall)	X	X	--	↑ *	AL↓*
Zelinski et al., 2014	Structural equation modeling	Age Sex Education	Time order judgment sound sweep discrimination task	X	X	→	--	--
			Recognition of syllables	↓*	X	↑ *	--	--

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. *SDM* Symbol-Digit Modalities. *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

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 setup 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 Roheger et al., 2020). 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). Furthermore, as women are at an increased risk of Alzheimer's disease (Scheyer et al., 2018), it could again be possible that they have more "room for improvement" at an earlier stage. This would again fit to the

compensation account (Lövden et al., 2012). 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 et al., 2018).

Three models found none of the investigated predictor to have a significant impact on changes after memory training when including among others age, sex, education, and 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 to 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 effects models) answer similar questions as the latent growth curve modeling 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 Mattes & Roheger, 2020). 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 by either 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 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.

Compliance with Ethical Standards

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., et al. (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.

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

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.

Lövdén, 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.

Mattes, A., & Roheger, M. (2020). Nothing Wrong About Change: The Adequate Choice of the Dependent Variable and Design in Prediction of Intervention Success. https://doi.org/10.21203/rs.3.rs-43704/v1.

McArdle, J. (1988). Dynamic but structural equation modeling of repeated measures data. In *Handbook of Multivariate Experimental Psychology* (pp. 561–614). Springer, Boston. 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., de Groot, J. A. H., Bouwmeester, W., Vergouwe, Y., Mallett, S. [. S.]., Altman, D. G., et al. (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., et al. (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.

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

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.*

Roozenbeek, B., Maas, A. I. R., Lingsma, H. F., Butcher, I., Lu, J., Marmarou, A., et al. (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. (2013). Within-cohort age-related differences in cognitive functioning. *Psychological science, 24*(2), 123–130.

Scheyer, O., Rahman, A., Hristov, H., Berkowitz, C., Isaacson, R. S., Brinton, R. D., & Mosconi, L. (2018). Female sex and Alzheimer's risk: the menopause connection. *The Journal of Prevention of Alzheimer's Disease, 5*(4), 225–230.

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., et al. (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., et al. (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., et al. (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.2014.0061.