Practice effects in performance outcome measures in patients living with neurologic disorders – A systematic review

Sven P. Holm, Arnaud M. Wolfer, Grégoire H.S. Pointeau, Florian Lipsmeier*, Michael Lindemann

Roche Pharma Research and Early Development, pRED Informatics, Roche Innovation Center Basel, F. Hoffmann-La Roche Ltd., Basel, Switzerland

ARTICLE INFO

Keywords:
Practice effects
Multiple sclerosis
Huntington disease
Parkinson disease
Dementia
Alzheimer disease
Mild cognitive impairment

ABSTRACT

Background: In this systematic review we sought to characterize practice effects on traditional in-clinic or digital performance outcome measures commonly used in one of four neurologic disease areas (multiple sclerosis; Huntington's disease; Parkinson's disease; and Alzheimer's disease, mild cognitive impairment and other forms of dementia), describe mitigation strategies to minimize their impact on data interpretation and identify gaps to be addressed in future work.

Methods: Fifty-eight original articles (49 from Embase and an additional 4 from PubMed and 5 from additional sources; cut-off date January 13, 2021) describing practice effects or their mitigation strategies were included.

Results: Practice effects observed in healthy volunteers do not always translate to patients living with neurologic disorders. Mitigation strategies include reliable changes indices that account for practice effects or a run-in period. While the former requires data from a reference sample showing similar practice effects, the latter requires a sufficient number of tests in the run-in period to reach steady-state performance. However, many studies only included 2 or 3 test administrations, which is insufficient to define the number of tests needed in a run-in period.

Discussion: Several gaps have been identified. In particular the assessment of practice effects on an individual patient level as well as the temporal dynamics of practice effects are largely unaddressed. Here, digital tests, which allow much higher testing frequency over prolonged periods of time, can be used in future work to gain a deeper understanding of practice effects and to develop new metrics for assessing and accounting for practice effects in clinical research and clinical trials.

1. Introduction

Chronic neurological diseases such as multiple sclerosis, Huntington's disease, Parkinson's disease or dementia may manifest in functional impairment in one or several functional domains (Lees et al., 2009; Roos, 2010; Sosnoff et al., 2014). Assessing these domains regularly can provide valuable insights into both the subject's disease status and the disease course and also inform treatment and disease management (Tur et al., 2018). Repeated performance assessments over time may, however, be susceptible to practice effects. Practice effects (also sometimes known as learning effects; see Panel for definition) is any change or improvement that results from repetition of tasks or activities, including repeated exposure to an instrument, rather than due to a true change in a patient's ability (Heilbroner et al., 2010; McCaffrey and Westervelt, 1995). For example, patients may perform better in subsequent tests as they fully comprehend the tasks (context memory or context effects) and gain knowledge of the sequence of tasks (episodic memory or content effects) and map the stimulus to the response (Goldberg et al., 2015). Over time this familiarity with the test could lead the subject to develop strategies that result in inflated test performance compared with a subject exposed to the test for the first time (Goldberg et al., 2015). The overall improvement in performance, or practice effects, is the result of consecutive gains that tend to be largest at first and gradually become smaller as the number of assessments increases (Figure 1) (Bartels et al., 2010; Falleti et al., 2006). In particular at short inter-test intervals, practice effects are often much greater than normative functional change over a similar interval (Jones, 2015).

Practice effects are often considered to introduce unwanted variance and thus complicate the interpretation of repeated clinical assessments (McCaffrey and Westervelt, 1995). If not accounted for, practice effects can lead to misdiagnosis or misinterpretation of clinical data, resulting in delayed access to the most effective treatment option (Elman et al., 2018;...
Marley et al., 2017). Despite a large body of literature on practice effects, their impact on the subject’s performance is seldom addressed and has been described as “large, pervasive and underappreciated” (Jones, 2015). Current study designs typically do not adequately estimate and mitigate their impact on test performance despite most repeated assessments being affected by practice effects to a varying degree (Johnson et al., 1991; McCaffrey and Westervelt, 1995). Thus, key in addressing this challenge is not only the characterization of practice effects and their underlying mechanisms, but also the implementation of mitigation strategies.

### 2. Methods

A systematic literature search was conducted on Embase and PubMed according to PRISMA guidelines to identify original articles that discuss practice effects in patients with neurologic disorders (cut-off date: January 13, 2021). Three separate search strings were used: one to identify original articles on commonly used performance outcome measures, one to identify original articles on practice effects, and one to identify original articles on either multiple sclerosis; Parkinson’s disease; Huntington’s disease; or mild cognitive impairment, Alzheimer’s disease or other forms of dementia (Table 1). Combining these three search strings with a Boolean “AND” resulted in the list of publications that were considered for this systematic review. Additional relevant records were identified through clinicaltrials.gov and from our own collection of references.

Publications were excluded if they were not original articles (for example, congress abstracts or other review articles); written in a language other than English, were duplicates; did not report practice effects or mitigation strategies in one of four disease areas specified in the search string; or did not report practice effects or mitigation strategies for one of the performance outcomes measures specified in the search string. ‘Neuropsychological test’ was included in the search string to identify original studies that investigated practice effects or mitigation strategies in test batteries that include at least one of the other performance outcome measures defined in the search string. This eligibility assessment was performed by the first author.

To minimize the impact of bias, only improvements in test performance that resulted from practice or repetition of task and cannot be

This systematic review aims to evaluate the presence and magnitude of practice effects associated with commonly used performance outcome measures in patients living with one of the four neurologic disorders: multiple sclerosis; Huntington’s disease; Parkinson’s disease; and Alzheimer’s disease, mild cognitive impairment and other forms of dementia.

In addition, this review discusses the different mitigation strategies that have been applied to minimize the impact of practice effects. Finally, it identifies gaps in our understanding of practice effects in patients with neurologic disorders, which should be addressed in future research.

### Figure 1. Schematic representation of the evolution of test performance through time solely due to task repetition (practice effects) when different assessment frequencies are considered. Each curve represents the test performance through time for a daily assessment (top curve) and a weekly assessment schedule (bottom curve). Each individual test is represented by a dot, colored either orange if it is part of the practice period (or run-in period), or blue if it is part of the steady-state period. During the practice period, performance gain between consecutive tests is largest at first and gradually reduces as the number of assessments increases. The assessment frequency does not alter the overall performance gain or number of iterations required to reach a steady-state suitable for reliable assessment, but decreases or increases the time needed to reach such state (e.g. 7 days vs 7 weeks). The subject’s abilities are considered constant over the period of time considered.

| Practice effects | Longitudinal effects | Run-in period | Iterations |
|------------------|----------------------|---------------|------------|
| Practice effects are any change or improvement that results from practice or repetition of task items or activities, including repeated exposure to an instrument, rather than due to a true change in an individual’s ability. Many studies, however, consider such improvements to be practice effects only if these improvements resulted in improved test scores. Practice effects are sometimes also known as learning effects. |
| Unlike practice effects, longitudinal effects describe changes in test performance resulting from functional changes, treatment intervention, or changes in motivation or fatigue levels. These longitudinal effects typically occur at larger timescales but may be confounded with practice effects. |
| A period/number of test iterations during which large practice effects are allowed to occur, until the magnitude of alterations from one test to the next is negligible. The run-in assessments are discarded and the subsequent test iteration is considered as a measure of baseline performance. Sometimes also known as ‘familiarization’ or ‘massed practice’ period. |
| Number of times a subject undertakes an assessment irrespective of the duration between test repetitions. |
explained by other means were considered to be practice effects. Thus, practice effects were considered whenever possible in the non-interventional cohort. Risk of publication bias and selective reporting was assessed by identifying the number of completed and potentially relevant studies listed on clinicaltrials.gov for which results have not been published yet.

In this systematic review, we aim to address the following five questions:

- Which metrics were used to identify possible practice effects?
- Were practice effects observed in patients with neurologic disorders, and how common were they?
- Which mitigation strategies were applied to minimize the impact of practice effects?
- Do practice effects carry any clinically meaningful information?
- Are there any gaps in our current understanding of practice effects in patients with neurologic disorders?

3. Results

The literature search on Embase and PubMed identified a total of 177 and 103 records, respectively. An additional 5 studies from a search on clinicaltrials.gov or from our own collection of references were included in the analysis. Of the 285 records, 58 were considered eligible (Figure 2).

In this systematic review, we aim to address the following five questions:

- Which metrics were used to identify possible practice effects?
- Were practice effects observed in patients with neurologic disorders, and how common were they?
- Which mitigation strategies were applied to minimize the impact of practice effects?
- Do practice effects carry any clinically meaningful information?
- Are there any gaps in our current understanding of practice effects in patients with neurologic disorders?

3.1. Identifying and quantifying practice effects

Several different approaches and metrics have been applied to identify practice effects, to quantify their magnitude and temporal dynamics, and to address potential biases in the interpretation of the data. These different approaches and metrics are summarized in Table S1 in the supplementary appendix.

3.1.1. Identifying practice effects

Descriptive statistics have been used to compare the change in performance between baseline and retest (Cohen et al., 2000, 2001; Duff et al., 2007, 2012; Duff and Hammers, 2022). However, it is more common to test the difference for statistical significance. Depending on the study design and the distribution of data, t-tests, Friedman’s test, Wilcoxon rank test, ANOVA, ANCOVA or other general linear models have been used to identify practice effects (Bachoud-Lévi et al., 2001; Barker-Collo, 2005; Beglinger et al., 2014a, 2014b; Benedict, 2005; Benedict et al., 2008; Benninger et al., 2011, 2012; Bever et al., 1995; Table 1.

| Search string 1: Performance outcome measures | Search string 2: Practice effects | Search string 3: Neurologic disorders |
|-----------------------------------------------|----------------------------------|-------------------------------------|
| Symbol Digit Modalities Test, Paced Auditory Serial Addition Test, Serial Reaction Time, Trail-Making Test, Stroop Test, Brief Visuospatial Memory Test-Revised, California Verbal Learning Test, Hopkins Verbal Learning Test | Practice Effects | Multiple Sclerosis |
| Upper extremity function: | Learning Effects | Parkinson’s Disease |
| Nine-Hole Peg Test, Pegboard Test, Speeded Tapping, | Retest Effects | Mild Cognitive Impairment, Alzheimer’s disease, Dementia |
| Gait & balance: | Initial Learning | Huntington’s Disease |
| Timed 25-Foot Walk, 2-Minute Walk Test, Timed Up and Go, Berg Balance Scale | | |
| Vision: | | |
| Composite scores: | | |
| Brief International Cognitive Assessment for Multiple Sclerosis (BICAMS), Consortium to Establish a Registry for Alzheimer’s Disease neuropsychological test battery (CERAD), Mini-Mental State Examination, Minimal Neuropsychological Assessment of MS Patients (MACFISM), Multiple Sclerosis Functional Composite (MSFC), Unified Huntington Disease Rating Scale (UHDRS), Wechslar Adult Intelligence Scale (WAIS), Wechsler Memory Scale (WMS), Neuropsychological Test | | |
| Digital performance outcome measures: | | |
| Digital, Computer, Tablet, Mobile, Smartphone | | |

While ‘neuropsychological test’ was included in the search string, this was only used to identify original articles that reported practice effects on at least one of the other performance outcome measures.

Figure 2. PRISMA flow diagram. Incl, Inclusion.
Table 2. Performance outcome measures and their functional domain.

| Performance outcome measure | Functional domain                             | Reference                                      |
|-----------------------------|----------------------------------------------|-----------------------------------------------|
| SDMT                        | Information processing speed, working memory | Smith (1982), Toh et al. (2014)               |
| PASAT                       | Information processing speed, working memory | Gronwall (1977), Rao et al. (1989)            |
| SRT                         | Sequential learning                          | Schendan et al. (2003)                       |
| TMT                         | Information processing speed                 | Saltz (1977)                                  |
| TMT-A                       | Information processing speed                 | Toh et al. (2014)                             |
| TMT-B                       | Executive function                           |                                               |
| Stroop Test                 | Information processing speed                 | Stroop (1935), Toh et al. (2014)             |
| BVMT-R                      | Visuospatial memory                          | Benedict (1997)                               |
| CVLT                        | Learning and memory                          | Delis et al. (1987), Elwood (1995)           |
| HVLT                        | Learning and memory                          | Brandt (1991)                                 |
| Coding/Digit Symbol         | Information processing speed                 | Wechsler (2008)                               |
| Digit Span                  | Working memory                               | Wechsler (2008)                               |
| Letter-Number Sequencing    | Working memory                               | Wechsler (2008)                               |
| Similarities                | Verbal comprehension                         | Wechsler (2008)                               |
| Matrix Reasoning            | Perceptual Organization                      | Wechsler (2008)                               |
| WMS                         | Working memory                               | Wechsler (2009)                               |
| Spatial Span                | Working memory                               | Wechsler (2009)                               |
| Logical Memory              | Episodic memory                              | Wechsler (2009)                               |
| Visual Reproduction         | Episodic memory                              | Wechsler (2009)                               |
| Paired Associations         | Verbal comprehension                         | Wechsler (2009)                               |
| MMSE                        | Global cognition                             | Folstein et al. (1975)                        |
| T25FW                       | Gait                                         | Moti et al. (2017)                            |
| 2MWIT                       | Gait                                         | Rossier and Wade (2001)                       |
| TUG                         | Gait, dynamic and static balance             | Podsiaido and Forchardson (1991)             |
| 9HPT                        | Hand-motor function, manual dexterity        | Feyes et al. (2017)                           |
| Purdue Pegboard             | Hand-motor function, manual dexterity        | Tiffin (1968)                                 |
| Speeded Tapping/Alternating Tapping | Hand-motor function, manual dexterity     | Stout et al. (2014), Prince et al. (2018), Westin et al. (2010) |
| Paced Tapping               | Hand-motor function, manual dexterity        | Stout et al. (2014)                           |
| Smartphone-based SDMT       | Information processing speed, working memory | Pham et al. (2021)                           |
| Memory Test                 | Short-term memory                            | Prince et al. (2018)                          |
| Brain on Track              |                                              |                                               |
| Attention task III          | Attention, information processing speed       | Ruano et al. (2020)                           |
| Visual memory task II       | Visual memory, attention                      | Ruano et al. (2020)                           |
| Delayed verbal memory       | Verbal memory                                | Ruano et al. (2020)                           |
| Calculus task               | Calculus                                     | Ruano et al. (2020)                           |
| Colour interference task    | Executive function                           | Ruano et al. (2020)                           |
| Verbal memory II            | Verbal memory                                | Ruano et al. (2020)                           |
| Opposite task               | Executive function, inhibitory control       | Ruano et al. (2020)                           |
| Written comprehension       | Language comprehension, information processing speed | Ruano et al. (2020)                           |
| Word categories             | Language                                     | Ruano et al. (2020)                           |
| Sequences                   | Executive function                           | Ruano et al. (2020)                           |
| Puzzles                     | Visuospatial abilities                       | Ruano et al. (2020)                           |
| CANTAB                      |                                              |                                               |
| One Touch Stockings of Cambridge | Executive function                   | Giedraitiene and Kubrys (2019)             |
| Spatial Working Memory      | Working memory                               | Giedraitiene and Kubrys (2019)               |
| Reaction Time Task          | Information processing speed                 | Giedraitiene and Kubrys (2019)               |
| Paired Associates Learning  | Visual memory                                | Giedraitiene and Kubrys (2019)               |
| MSReactor                   |                                              |                                               |
| Simple Reaction Time        | Information processing speed                 | Merlo et al. (2019)                           |
| Choice Reaction Time        | Visual attention                             | Merlo et al. (2019)                           |
| One-Back Test               | Working memory                               | Merlo et al. (2019)                           |

(continued on next page)
Practice effects may also be indirectly identified. For example, a change in clinical diagnosis (for example, from mild cognitive impairment to cognitively healthy) can indicate the presence of practice effects (Duff et al., 2011). Similarly, patients who showed functional disability at baseline may, due to practice, improve their performance at retest sufficiently to no longer be classified as impaired (Schwid et al., 2007). Practice effects may also be indirectly inferred from a reliable change index analysis (Rosas et al., 2020). As the reliable change index captures the expected change based on the change observed in a reference population. An improvement beyond this index suggests that the patient showed greater than expected improvements, which may indicate practice effects.

For the purpose of this review, any improvements in test performance that cannot be explained by other means such as interventional effects, functional recovery or decline etc. were considered to be practice effects (Merlo et al., 2019).

Practice effects may also be indirectly identified. For example, a change in clinical diagnosis (for example, from mild cognitive impairment to cognitively healthy) can indicate the presence of practice effects (Duff et al., 2011). Similarly, patients who showed functional disability at baseline may, due to practice, improve their performance at retest sufficiently to no longer be classified as impaired (Schwid et al., 2007). Practice effects may also be indirectly inferred from a reliable change index analysis (Rosas et al., 2020). As the reliable change index captures the expected change based on the change observed in a reference population. An improvement beyond this index suggests that the patient showed greater than expected improvements, which may indicate practice effects.

For the purpose of this review, any improvements in test performance that cannot be explained by other means such as interventional effects, functional recovery or decline etc. were considered to be practice effects.

### 3.1.3. Estimating the temporal dynamics of practice effects

Besides quantifying the magnitude of practice effects, few studies have also estimated the duration until steady-state performance is reached. In Prince et al. (2018), the steady-state index was computed as the first confirmed test iteration at which the performance reached a predefined threshold. In contrast, Pham et al. (2021) fitted a biphasic, linear + linear learning curve model to the data, with the first phase fitting the practice phase and the second phase the steady-state performance phase. Using non-linear regression, they identified the change point, which marked the end of the practice phase.

### 3.1.4. Addressing biases

Finally, few analyses attempted to account for various biases. These include accounting for the attrition effect (Elman et al., 2018), which describes the bias introduced by patients lost to follow-up, and for the reduced capacity for practice effects among patients with high test performance at baseline (Gross et al., 2018; Sormani et al., 2019).

### 3.2. Practice effects

Across all four disease areas, certain performance outcome measures, or tests, were more prone to practice effects than others (Table 3). Among those assessing information processing speed, the Paced Auditory Serial Addition Test (PASAT) was most strongly associated with practice effects. A possible explanation is its stronger working memory component

---

**Table 2 (continued)**

| Performance outcome measure | Functional domain | Reference |
|-----------------------------|-------------------|-----------|
| Detection Task              | Information processing speed | Hammers et al. (2011) |
| Identification Task         | Visual attention   | Hammers et al. (2011) |
| One-Back Task               | Working memory     | Hammers et al. (2011) |
| One Card Learning           | Visual recognition | Hammers et al. (2011) |
| Divided Attention           | Divided attention  | Hammers et al. (2011) |
| Associative Learning        | Associative learning | Hammers et al. (2011) |
| Visual Search               | Cognitive function, motor behavior | Utz et al. (2013) |
| MSPT                         |                    |           |
| Manual Dexterity Test       | Hand-motor function, manual dexterity | Rao et al. (2020) |
| Contrast Sensitivity Test   | Vision             | Rao et al. (2020) |
| Walking Speed Test          | Gait               | Rao et al. (2020) |
| Driving Simulator           | Visual information integration | Teasdale et al. (2016) |

2MWT, Two-Minute Walk Test; 9HPT, Nine-Hole Peg Test; BVMT-R, Brief Visuospatial Memory Test-Revised; CANTAB, Cambridge Neuropsychological Test Automated Battery; CVLT, California Verbal Learning Test; HVLT, Hopkins Verbal Learning Test; MMSE, Mini-Mental State Examination; MSPT, Multiple Sclerosis Performance Test; PASAT, Paced Auditory Serial Addition Test; SDMT, Symbol Digit Modalities Test; SRT, Serial Reaction Time; T25FW, Timed 25-Foot Walk; TMT, Trail-Making Test; TUG, Timed Up and Go; WAIS, Wechsler Adult Intelligence Scale; WMS, Wechsler Memory Scale.
and its increased complexity. Among tests of executive function, the Trail-Making Test B (TMT-B) was more likely to produce practice effects than the Stroop Interference Test. The inhibitory processes involved when performing the Stroop Interference Test might make this test less prone to practice effects. On tests of learning and memory or visuospatial memory such as the Hopkins Verbal Learning Test (HVLT) or the Brief Visuospatial Memory Test-Revised (BVMT-R), practice effects were more common if the same form was used. This suggest that practice effects are mostly driven by item learning. In addition, on the BVMT-R, delayed recall was more often associated with practice effects than immediate recall. However, this was not observed on the HVLT.

3.3. Practice effects in patients with multiple sclerosis

The literature search revealed 27 studies on practice effects in multiple sclerosis. Details on study design and presence of practice effects are summarized in Table 4. Effect sizes (Cohen’s d) are depicted in Figure 3.

Repeated assessment of information processing speed was likely to result in practice effects. Both the traditional, clinician-administered as well as the smartphone-based Symbol Digit Modalities Test (SDMT) produced practice effects in most studies, although they were minimal and smaller than observed in healthy controls (Cohen’s d: 0.2 vs 0.8) (Benedict, 2005; Benedict et al., 2008; Eshaghi et al., 2012; Glanz et al., 2012; Pham et al., 2021; Reilly and Hynes, 2018; Schiwid et al., 2007; Vogt et al., 2009). Only Fuchs et al. (2020) and Bever et al. (1995) noted an absence of practice effects. However, all patients were previously exposed to the SDMT prior to enrolment, which might have impacted the ability to detect practice effects (Fuchs et al., 2020). Practice effects were also common on the PASAT, both with short inter-test intervals (every two weeks or shorter) and long inter-test intervals (every month or longer) (Barker-Collo, 2005; Benedict, 2005; Bever et al., 1995; Cohen et al., 2000, 2001; Eshaghi et al., 2012; Glanz et al., 2012; Nagels et al., 2008; Rosti-Otajärvi et al., 2008; Schiwid et al., 2007; Solari et al., 2005; Sormani et al., 2019; Utz et al., 2016). Compared with the SDMT, PASAT-related practice effects were larger in magnitude (Cohen’s d: 0.3–0.4 vs 0.2) (Benedict, 2005; Eshaghi et al., 2012). A trend towards improved test scores was also noted on the TMT-A in Reilly and Hynes (2018); however, all patients underwent cognitive rehabilitation prior to retest. Practice effects were also discernable on the Word, but not Color subtest of the Stroop Test (Pliskin et al., 1996).

There was little evidence of practice effects associated with executive function. On the TMT-B, subtle practice effects were observed in Reilly and Hynes (2018), but not in Pliskin et al. (1996). By comparison, repeated testing with the Stroop Interference test did not result in practice effects (Pliskin et al., 1996).

Practice effects were common on assessments of learning and memory or visuospatial memory. Both the California Verbal Learning Test (CVLT) and the BVMT-R produced practice effects. This was particularly evident if the same form was used (Benedict, 2005; Eshaghi et al., 2012). On the Visual Reproduction test, improvement in performance was independent of treatment allocation only on immediate recall but not on delayed recall (Pliskin et al., 1996). On the latter, improved test scores were observed only in patients receiving high-dose interferon-β. Finally, the digital Visual Search test was not associated with practice effects (Utz et al., 2015).

On the Digit Span test, a measure of working memory, improved test scores were only observed in the backward condition (Vogt et al., 2009). However, this improvement was associated with additional cognitive training.

On digital cognitive batteries, practice effects were observed on the Brain on Track test (Ruano et al., 2020), the MSReactor (Merlo et al., 2019) and the Cambridge Neuropsychological Test Automated Battery (CANTAB) (Giedraitiene and Kaubrys, 2019), with larger practice effects associated with more demanding tasks (Giedraitiene and Kaubrys, 2019; Merlo et al., 2019).

On gait and balance tests, short-term practice effects were reported on the Timed Up and Go (Meyer et al., 2020). In contrast, most studies showed an absence of practice effects on the Timed 25-Foot Walk (T25FW) (Cohen et al., 2000, 2001; Patzold et al., 2002; Rosti-Otajärvi et al., 2008; Solari et al., 2005), Two-Minute Walk Test (2MWT) (Meyer et al., 2020) and the digital Walking Speed Test (Rao et al., 2020); only
| Study                          | Sample size | Study type | Follow-up duration | Practice effects in cohort of interest | Comment                                                                 |
|-------------------------------|-------------|------------|--------------------|----------------------------------------|-------------------------------------------------------------------------|
| Barker-Collo (2005)           | MS: 30      | LO         | Single session     | PASAT                                  | Practice effects on the PASAT were observed for the 2.0-, 1.6-, and 1.2-second presentation, but not for the 2.4-second presentation. |
| Benedict et al. (2005)        | MS: 34      | LO         | 1 week             | SDMT                                   | Practice effects on the BVMT-R and CVLT were observed only with the same form. |
| Benedict et al. (2008)        | MS: 85      | LO         | 5 months           | SDMT                                   | An ANOVA was conducted to investigate the main effect over time among patients with multiple sclerosis. |
| Bever et al. (1995)           | MS: 19      | RCT        | 16 weeks           | PASAT, SDMT                            | All patients randomized to the active treatment arm had been off the study drug (3,4-diaminopyridine) for at least 30 days at the time of each evaluation. |
| Cohen et al. (2000)           | MS: 10      | LO         | 6 months           | PASAT, T25FW                           | Practice effects were assessed during a run-in period prior to randomization. |
| Cohen et al. (2001)           | MS: 436     | RCT        | 28 days            | PASAT, 9HPT, T25FW                     | Results are reported for a composite score. |
| Erlanger et al. (2014)        | MS: 59      | LO         | 45 days            | SDMT                                   |                                                                         |

(continued on next page)
| Study | Sample size | Cohort of interest | Study type | Follow-up duration | # test iteration | Practice effects in cohort of interest | Comment |
|-------|-------------|--------------------|------------|-------------------|-----------------|----------------------------------------|---------|
| Eshaghi et al. (2012) | MS: 41 | Add. cohort | Continuous effects | Mean (SD) of 10.8 (3.78) days | 2 | PASAT, SDMT, BVMT-R (total recall) | A total of 158 patients were recruited, of which 41 were included in the practice effects analysis. A trend towards improvement was observed on the SDMT. |
| Fuchs et al. (2020) | MS: 531 | LO | Initial effects | 16 years | ≤10 | SDMT, BVMT-R (delayed recall), CVLT (total and delayed recall) | |
| Gallus and Mathiowetz (2003) | MS: 35 | LO | Initial effects | 1 week | 2 | Purdue Pegboard: Sum of three trials (bimanual), Purdue Pegboard: One trial (dominant hand, non-dominant hand, bimanual, assembly) | |
| Giedraitiene and Kubrys (2019) | Relapsing MS: 30, Stable MS: 30 | LO | Initial effects | 3 months | 3 | CANTAB: One Touch Stockings of Cambridge, CANTAB: Spatial Working Memory | Practice effects were only assessed in patients with relapsing MS. Functional recovery and practice effects may have occurred concurrently in relapsing MS. |
| Glanz et al. (2012) | MS: 69 | LO | Initial effects | 5 years | 7 | PASAT, SDMT | |
| Merlo et al. (2019) | MS: 328, HC: 30 | LO | Initial effects | 18 months | ≤10 | MSReactor: Simple Reaction Time, Choice Reaction Time, One Back | A total of 450 patients with MS were recruited and completed their baseline assessment, practice effects were assessed in a subset of 328 patients with MS who completed up to 10 assessments. |
| Study            | Sample size | Add. cohort | Study type | Follow-up duration | # test iteration | Practice effects in cohort of interest | Comment                                                                 |
|-----------------|-------------|-------------|------------|-------------------|------------------|----------------------------------------|--------------------------------------------------------------------------|
| Meyer et al. (2020) | MS: 10      | HC: 40      | LO         | 4-5 weeks         | 4                | T25FW                                  | Practice effects are reported for 8 patients with MS; 2 patients with MS were excluded due to muscle exhaustion/pain. |
| Nagels et al. (2008) | MS: 110     | LO          | Single session | 2                 |                  | TUG                                    |                                                                         |
| Patzold et al. (2002) | MS untreated controls: 10 | MS receiving steroid therapy for acute relapse: 27 | NRI         | 20 days           | 3                | PASAT                                  |                                                                         |
| Pham et al. (2021)  | MS: 15      | LO          | ≥20 weeks   | ≥20               |                  | Smartphone-based SDMT                  | A total of 154 patients and 39 healthy controls were recruited, of which 15 patients and 1 healthy control were included in the practice effects analysis. |
| Pliskin et al. (1996) b | MS with high-dose IFN-β-1b: 9 | RCT         | 2 years     | 2                 |                  | Stroop Word Test, WMS: Visual Reproduction (delayed recall), TMT-B | Main effect of time was observed for improvement on Stroop Word Test and WMS Visual Reproduction (immediate recall); improvement on WMS Visual Reproduction (delayed recall) was associated with high-dose IFN-β-1b. |
| Rao et al. (2020)   | MS: 30      | HC: 30      | LO         | Single session    | 2                | MSPT: Manual Dexterity Test, Stroop Interference Test | WMS: Logical Memory, Purdue Pegboard, MSPT Walking Speed Test |
| Study | Sample size | Study type | Follow-up duration | # test iteration | Practice effects in cohort of interest | Comment |
|-------|-------------|------------|--------------------|------------------|----------------------------------------|---------|
|       | Cohort of interest | Add. cohort |                   |                  | Continuous effects | Initial effects | Inconclusive | No improvement | |
| Reilly and Hynes (2018) | MS receiving cognitive rehabilitation: 12 | NRI | 18 weeks | 3 | SDMT | BVMT-R (delayed recall) | Observed improvements on were associated with cognitive rehabilitation; improvements on the SDMT and the TMT-A did not reach statistical significance. |
|       | | | | | | | | | |
| | | | | | | | | |
| Rosti-Otajärvi et al. (2008) | MS: 10 | HC: 10 | LO | 4 weeks | 5 | PASAT | T25FW | |
| | | | | | | | | |
| Ruano et al. (2020) | MS: 30 | HC: 30 | LO | 3 months | 4 | Brain on Track: Opposite Task | Brain on Track: Attention III | Brain on Track: Delayed Verbal Memory |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| Schwid et al. (2007) | MS: 153 (pooled analysis of 74 patients initially allocated to placebo and 79 patients initially allocated to glatiramer acetate) | RCT with OLE | 10 years | 3 | SDMT | | A total of 251 patients were initially randomized, of whom 153 had 10-year follow-up data and were included in the analyses. Improvements at year 2 were independent of initial treatment allocation. |
| | | | | | | | | |
| | | | | | | | | |

(continued on next page)
| Study                                     | Sample size | Study type | Follow-up duration | # test iteration | Practice effects in cohort of interest | Comment                      |
|------------------------------------------|-------------|------------|--------------------|------------------|---------------------------------------|------------------------------|
|                                          | Cohort of interest | Add. cohort |                    |                  | Continuous effects | Initial effects | Inconclusive | No improvement |                      |
| Solari et al. (2005)                     | MS: 32      | LO         | Single session     | 6                | PASAT                  | T25FW            |              |                |                      |
|                                          |             |            |                    |                  |                        |                 |              |                |                      |
| Sormani et al. (2019)                    | MS: 1,009 (randomized 1:1:1 to receive either fingolimod 0.5 or 1.25 mg once daily or placebo) | RCT | 2 weeks | 3 | PASAT | Practice effects were assessed during a run-in period prior to randomization. |
| Utz et al. (2016)                        | MS: 44 (pooled analysis of 22 patients receiving fingolimod, 11 natalizumab, 7 interferon and 1 glatiramer acetate) | NRI | 1 year | 3 | PASAT | Initially 73 patients were recruited, of whom 41 had follow-up data and did not switch therapy. |
|                                          |             |            |                    |                  |                        |                 |              |                |                      |
| Vogt et al. (2009)                       | MS with high-intensity cognitive training: 15 | NRI | 4–8 weeks | 3 | SDMT | WMS: Digit Span |
|                                          |             |            |                    |                  |                          |                 | WMS: Spatial Span |
|                                          |             |            |                    |                  |                          |                 | WMS: Logical Memory |
|                                          |             |            |                    |                  |                          |                 | Visual Search |
|                                          | MS with distributed training: 15 |             |                    |                  | PASAT | WMS: Digit Span (forward) |
|                                          | MS controls: 15 |             |                    |                  |                        |                 | WMS: Digit Span (backwards) |

9HPT, Nine-Hole Peg Test; Add., additional; Approx., approximately; BVMT-R, Brief Visuospatial Memory Test-Revised; CANTAB, Cambridge Neuropsychological Test Automated Battery; CVLT, California Verbal Learning Test; HC, healthy controls; MS, multiple sclerosis; MSPT, Multiple Sclerosis Performance Test; PASAT, Paced Auditory Serial Addition Test; RCT, randomized controlled trial; SDMT, Symbol Digit Modalities Test; T25FW, Timed 25-Foot Walk; TMT-A/B, Trail-Making Test A/B; TUG, Timed Up and Go; WMS, Wechsler Memory Scale.

* Continuous effects’ is defined as a continuous improvement in test performance for ≥4 test administrations, with test performance continuing to improve up to the last test administered. By definition this can only be applied to studies that administered the test at least 4 times. In all other instances, practice effects are described as either initial effect if clear signs of practice effects were evident; inconclusive if practice effects were observed for a selection of test metrics, in a subgroup of patients only, or if other reasons may explain the improvement in test performance (for example, due to contribution of other tests in composite scores, or association with additional training or treatment etc.); or no effect if no improvement in test performance was observed.

b Results of the repeated assessments were not consistently reported for the placebo cohort; hence outcomes for the total cohort are reported.
Meyer et al. (2020) demonstrated discernable practice effects on the T25FW. On assessments of hand-motor function, practice effects were observed on the digital Manual Dexterity Test (Rao et al., 2020) and Nine-Hole Peg Test (9HPT) (Cohen et al., 2000, 2001; Rosti-Otajärvi et al., 2008; Solari et al., 2005). However, in Patzold et al. (2002), 9HPT-related improvements were only observed in those patients receiving active treatment for acute relapse. By comparison, practice effects were unlikely to occur on the Purdue Pegboard, especially when each hand was considered separately (Gallus and Mathiowetz, 2003; Pliskin et al., 1996).

Finally, there was no evidence of practice effects on the Contrast Vision Test (Rao et al., 2020).

3.4. Practice effects in patients with Parkinson’s disease

The literature search revealed seven studies on practice effects in Parkinson’s disease. Details on study design and presence of practice effects are summarized in Table 5. Only one study reported effect sizes (Cohen’s d), which are depicted in Figure 4.

Most of these studies did not reveal any practice effects, whether on the CVLT (Higginson et al., 2009), the Digit Span Test (Turner et al., 2016), the Similarities Test (Turner et al., 2016) or the digital Tapping Test (Westin et al., 2010). On the Serial Reaction Time test, two studies revealed improvements, or reduced reaction times, that suggest practice effects (Benninger et al., 2011, 2012). However, Buelow et al. (2015) noted a worsening at retest.

3.5. Practice effects in patients with Huntington’s disease

The literature search revealed seven studies on practice effects in Huntington’s disease. Details on study design and presence of practice effects are summarized in Table 6. Only one study reported effect sizes (Cohen’s d), which are depicting in Figure 5.

The practice effects analyses revealed mostly mixed results. On the SDMT, for example, practice effects were observed in two studies, with patients with pre-manifest Huntington’s disease showing larger practice effects than patients with manifest Huntington’s disease (Beglinger et al., 2014a; Stout et al., 2014). However, one study did not find any discernable practice effects (Duff et al., 2007). On the TMT-A, larger practice effects were observed in patients with manifest Huntington’s disease as opposed to pre-manifest Huntington’s disease (Stout et al., 2014). Mixed results were also obtained on the Stroop Word Test (Beglinger et al., 2014a; Snowden et al., 2001; Stout et al., 2014) and the Stroop Color Test (Beglinger et al., 2014a; Snowden et al., 2001).

By comparison, the initial improvement on the TMT-B observed within 1–3 days was of similar magnitude in patients with pre-manifest or with manifest Huntington’s disease (Stout et al., 2014). When assessed annually, the initial gain was followed by a decline in
Table 5. Practice effects in patients with Parkinson’s disease.

| Study | Sample size | Practice effects in cohort of interest | Comment |
|-------|-------------|---------------------------------------|---------|
| | | Continuous effects | Initial effects | Inconclusive | No improvement |
| Benninger et al. (2011) | PD receiving sham intervention: 13 | Serial Reaction Time | Practice effects were independent of the intervention. |
| | PD receiving iTBS: 13 | | |
| Benninger et al. (2012) | PD receiving sham intervention: 13 | Serial Reaction Time | Practice effects were independent of the intervention. |
| | PD receiving rTMS: 13 | | |
| Buelow et al. (2015) | PD receiving placebo: 20 | Serial Reaction Time | Practice effects were only assessed in the control cohort. |
| | PD receiving galantamine hydrobromide ER: 33 | | |
| Higginson et al. (2009) | PD: 22 | CVLT (total recall, delayed recall) | |
| Prince et al. (2018) | PD: 312 (Tapping test) | Tapping test | |
| | YHC: 150 (Tapping test); 10 (Memory test) | ≥20 Tapping tests; ≥20 memory tests | |
| | LO: 6 months | | |
| Turner et al. (2016) | PD with MCI receiving placebo: 15 | WAIS: Similarities test | Practice effects were only assessed in the placebo cohort. |
| | PD with MCI receiving atomoxetine: 15 | | |
| Westin et al. (2010) | PD receiving duodenal levodopa/carbidopa: 65 | Hand Computer Tapping Test | No difference was observed over the first three days and remaining days. |

Add., additional; CVLT, California Verbal Learning Test; HC, healthy controls; iTBS, intermittent theta-burst stimulation; LO, longitudinal observational; MCI, mild cognitive impairment; NRI, non-randomized interventional; PD, Parkinson’s disease; RCT, randomized controlled trial; rTMS, repetitive transcranial magnetic stimulation; SDMT, Symbol Digit Modalities Test; WAIS, Wechsler Adult Intelligence Scale; YHC, young healthy controls.

* Continuous effects is defined as a continuous improvement in test performance for ≥4 test administrations, with test performance continuing to improve up to the last test administered. By definition this can only be applied to studies that administered the test at least 4 times. In all other instances, practice effects are described as either ‘initial effect’ if clear signs of practice effects were evident; ‘inconclusive’ if practice effects were observed for a selection of test metrics, in a subgroup of patients only, or if other reasons may explain the improvement in test performance (contribution of other tests in composite scores, association with additional training or treatment etc.); or ‘no effect’ if no improvement in test performance was observed.

Add., additional; CVLT, California Verbal Learning Test; HC, healthy controls; iTBS, intermittent theta-burst stimulation; LO, longitudinal observational; MCI, mild cognitive impairment; NRI, non-randomized interventional; PD, Parkinson’s disease; RCT, randomized controlled trial; rTMS, repetitive transcranial magnetic stimulation; SDMT, Symbol Digit Modalities Test; WAIS, Wechsler Adult Intelligence Scale; YHC, young healthy controls.

* Continuous effects is defined as a continuous improvement in test performance for ≥4 test administrations, with test performance continuing to improve up to the last test administered. By definition this can only be applied to studies that administered the test at least 4 times. In all other instances, practice effects are described as either ‘initial effect’ if clear signs of practice effects were evident; ‘inconclusive’ if practice effects were observed for a selection of test metrics, in a subgroup of patients only, or if other reasons may explain the improvement in test performance (contribution of other tests in composite scores, association with additional training or treatment etc.); or ‘no effect’ if no improvement in test performance was observed.
3.6. Practice effects in patients with mild cognitive impairment, Alzheimer's disease or other forms of dementia

The literature search revealed 18 studies on practice effects analyses in mild cognitive impairment, Alzheimer's disease and other forms of dementia. Details on study design and presence of practice effects are summarized in Table 7, and effects sizes (Cohen's $d$) are depicted in Figure 6.

Compared with the other disease areas, practice effects were less common on test of information processing speed, especially on the SDMT (Duff et al., 2017, 2018; Duff and Hammers, 2022) and TMT-A (Duff et al., 2017, 2018; Duff and Hammers, 2022). However, when present, they tend to be smaller than observed in healthy controls (Duff et al., 2015), and their magnitude correlated significantly with hippocampal volume ($r = 0.73; P < 0.01$) (Duff et al., 2018). Practice effects on the SDMT and the Word and Color subtests of the Stroop Test were more likely to be observed in patients with greater levels of cognitive impairment (Rosas et al., 2020). Moreover, a trend towards improved test scores were observed on the Stroop Word test and subtle practice effects on the Stroop Color test in a mixed cohort of patients with mild cognitive impairment and healthy volunteers even after a long inter-test interval of six years (Elman et al., 2018). Practice effects on the Digit-Symbol or Coding Test were also more likely to occur with increasing levels of cognitive impairment (Rosas et al., 2020). However, Duff et al. (2012) reported an inverse correlation between the magnitude of practice effects and dementia severity measured by MMSE (partial $r = 0.26; P = 0.046$; Cohen’s $d = 0.54$), even after controlling for baseline performance.

For tests assessing executive function such as the TMT-B or the Interference subtest of the Stroop Test, there was little evidence of practice effects (Britt et al., 2011; Duff et al., 2015, 2017, 2018; Duff and Hammers, 2022; Elman et al., 2018). Nonetheless, practice effects were more likely to occur with increasing levels of cognitive impairment (Rosas et al., 2020) or in specific subgroups (Frank et al., 1996).

Repeated testing with the CVLT, a measure of learning and memory, resulted in practice effects, especially on less demanding tasks such as short delayed free or cued recall and long delayed cued recall (Campos-Magdaleno et al., 2017; Elman et al., 2018). The lack of practice effects on the more memory-demanding tasks of the CVLT, including long delayed free recall, suggests that explicit memory deteriorates in amnestic mild cognitive impairment while implicit memory involved in practice effects is still preserved (Campos-Magdaleno et al., 2017). Practice effects were also observed on both total and delayed recall of the HVLT when retested within a week (Duff et al., 2017, 2018). In patients with probable Alzheimer's disease, stronger practice effects correlated inversely with disease severity measured by MMSE after controlling for baseline performance (partial $r = 0.47; P < 0.001$; Cohen's $d = 1.016$) (Duff et al., 2012).

On the Visual Reproduction test, improvements suggestive of practice effects were observed in patients at risk of developing Alzheimer’s disease for both delayed and immediate recall, while the performance on both tasks tended to remain stable or worsen in patients diagnosed with Alzheimer's disease (Frank et al., 1996). By comparison, a mixed cohort of patients with mild cognitive impairment and healthy volunteers showed definite practice effects only on delayed recall (Elman et al., 2018). Finally, on the Logical Memory test, practice effects were more common in patients with mild cognitive impairment than in patient with Alzheimer's disease (Britt et al., 2011; Claus et al., 1993; Elman et al., 2018; Gavett et al., 2016).

On the BVMT-R, a measure of visuospatial memory, practice effects were reported on both total and delayed recall, in particular with short inter-test intervals (Duff et al., 2007, 2015, 2017, 2018). However, there was little-to-no signs of practice effects if the inter-test interval was increased to one year or longer (Duff and Hammers, 2022). Furthermore, the magnitude of practice effects on delayed recall correlated with $^{18}$F-flutemetamol uptake in amyloid plaques ($r = -0.45; P = 0.02$; Cohen's $d = 0.71$) (Duff et al., 2014).

Working memory was assessed with a couple of different tests. Digit Span, in particular in the backward condition, Spatial Span and Letter-Number Sequencing were all associated with practice effects (Elman et al., 2018). Similarly, repeated testing withCogState's One-Back Test resulted in reduced reaction times in patients with either mild cognitive impairment or dementia with Lewis Bodies and in improved accuracy scores in the entire study cohort (Hammers et al., 2011).
### Table 6. Practice effects in patients with Huntington's disease.

| Study            | Sample size | Study type | Follow-up duration | # test iterations | Practice effects in cohort of interest | Comment                                                                 |
|------------------|-------------|------------|--------------------|-------------------|----------------------------------------|-------------------------------------------------------------------------|
| Bachoud-Lévi et al. (2001) | HD: 22      | LO         | 2–4 years          | 3–5               | TMT-B                                  |                                                                         |
| Beglinger et al. (2014a) | HD: 34 (randomized 1:1 to receive citalopram or placebo) | RCT | ≤24 hours and ≥6 days (mean [SE]: 20.4 [2.2] days) | 2                 | SDMT                                   | Practice effects were assessed prior to randomization. Initial effects on the SDMT were observed only with longer inter-test interval. |
| Beglinger et al. (2014b) | HD receiving placebo: 15 | HD receiving 20 mg citalopram: 16 | RCT | 20 weeks              | 6                                      | Stroop Word Test                                                       |
| Duff et al. (2007)   | HD: 170     | LO         | Mean (SD) of 220 (122) days | 2                 | SDMT                                   |                                                                         |
| Snowden et al. (2001) | HD: 87      | LO         | 1–3 years          | 2–4               | Stroop Interference Test                |                                                                         |
| Stout et al. (2014) | HD: 56      | LO         | 5–7 weeks          | 3                 | SDMT                                   | Practice effects were only assessed in patients with HD or pre-HD, but not in HC. Practice effects on the TMT-A were observed only in HD patients, while practice effects on the Speed Tapping and Paced Tapping tests were only observed in pre-HD, patients. |

(continued on next page)
| Study Sample size | Study type | Practice effects in cohort of interest | Comment |
|------------------|------------|---------------------------------------|---------|
| Toh et al. (2014) | HD: 22/C15 | SDMT, Stroop Word Test, Stroop Color Test, Stroop Interference Test, BVMT-R, CVLT | No improvement |
|                   | HC: 22 LO 12 months | TMT-A/C15, TMT-B/C15, MMSE | No improvement |
|                   |                       | WAIS: Digit Span | |
computed from the difference between the predicted and observed scores at retest can be used to define a threshold for detecting functional decline or functional recovery beyond the expected practice effect.

### 3.7.3. Replacement method

The replacement method of Elman et al. (2018) estimates group-level, attrition-corrected practice effects. With this method, the cohort at retest (i.e., the returnee cohort) is compared against a test-naïve, age-matched cohort (i.e., the replacement cohort). Any difference observed between these two cohorts is assumed to be a combination of attrition and practice effects. Attrition-corrected practice effects are obtained by subtracting the difference in mean scores of the overall cohort at baseline (i.e., mean baseline scores of returnees and those lost to follow-up) and the returnee cohort at baseline (attrition effect) from the difference in mean scores of a separate, test-naïve replacement cohort at baseline and the returnee cohort retest (difference score). These estimated practice effects can then be subtracted from the test scores of the returnee cohort obtained at retest, resulting in practice-effect-corrected retest scores. This methodology is more robust for larger sample size of the overall cohort and the cohort lost to follow-up. Depending on the drop-out rate, the replacement cohort can be small (and thus returnee population large), resulting in instability in the calculation of the difference score which is a key part of the attrition-corrected practice effect value. Furthermore, this methodology has been demonstrated for a single retest. While it is possible to apply it to more than one retest, it would require the management of multiple cohorts as retests (i.e., multiple replacement and returnee cohorts). Finally, data from a test-naïve, age-matched replacement cohort may not always be available for less established tests.

### 3.7.4. Alternative forms

Few studies purposely administered the same form to maximize practice effects (Duff et al., 2011, 2017, 2018). Conversely, the use of alternative forms – if available – can help reduce practice effects if they are driven by learning a specific sequence of items (Beglinger et al., 2014b; Benedict, 2005). In fact, several studies reported on absence of practice effects when using the alternative form, including on the SDMT (Fuchs et al., 2020), the CVLT (Eshaghi et al., 2012) or the Wechsler Memory Scale (Claus et al., 1991).

Moreover, in a direct comparison, the use of an alternative form prevented practice effects on all CVLT and BVMT-R metrics (Benedict, 2005). This is contrary to the practice effects observed on both measures in Eshaghi et al. (2012) and in Reilly and Hynes (2018), suggesting that alternative forms may only reduce but not fully prevent practice effects. Mixed results were also obtained on the HVLT, where only patients with pre-manifest, but not manifest, Huntington’s disease showed signs of practice effects on the alternative form (Stout et al., 2014).

Consistent with the literature (Bever et al., 1995; Cohen et al., 2000, 2001; Eshaghi et al., 2012; Glanz et al., 2012; Nagels et al., 2008; Rosti-Otajärvi et al., 2008), the direct comparison of Benedict (2005) revealed practice effects on both SDMT (group by time interaction effect: \( P > 0.05 \)) and PASAT (Cohen’s d for same form: 0.3; for alternative form: 0.4), irrespective whether the same and alternative form was used. This suggests that patients may develop over time more effective test taking strategies and that this drives the practice effects seen despite the use of alternative forms (Beglinger et al., 2014a; Gross et al., 2018). Consequently, other strategies are needed to minimize the impact of practice effects.

---

**Figure 5.** Repeated-measures effect sizes of the observed changes between test iterations in Huntington’s disease obtained from Stout et al. (2014). Light green dots (○) indicate initial practice effects. Red dots (●) indicate an absence of practice effects, as defined in Table 6. * Effect size reported for the change observed between the second and third test iteration rather than between the first and second test iteration. CVLT, California Verbal Learning Test; HD, Huntington’s disease; HVLT, Hopkins Verbal Learning Test; pre-HD, pre-manifest Huntington’s disease; SDMT, Symbol Digit Modalities Test; TMT, Trail-Making Test.
Table 7. Practice effects in patients with either mild cognitive impairment, Alzheimer's disease or other forms of dementia.

| Study                           | Sample size | Study type | Follow-up duration | # test iterations | Practice effects in cohort of interest | Comment |
|---------------------------------|-------------|------------|--------------------|-------------------|---------------------------------------|---------|
| Britt et al. (2011)             | MCI: 48     | LO 60 months | 2-8               |                   | Continuous effects                     | TMT-B   |
|                                 | HC: 36      |            |                    |                   | Initial effects                        |         |
|                                 | AD: 28      |            |                    |                   | Inconclusive                           |         |
|                                 |             |            |                    |                   | No improvement                         |         |
| Campos-Magdaleno et al. (2017)  | MCI: 21     | LO 18 months | 2                 |                   | CVLT (total recall, short delayed free recall, short delayed cued recall, long delayed cued recall, long delayed free recall) | Practice effects for total recall and long delayed free recall were observed in only patients with SMC, but not in patients with MCI. |
| Claus et al. (1991)             | AD: 17      | LO 2 weeks  | 3                 |                   | WMS: Logical Memory                    |         |
|                                 | HC: 16      |            |                    |                   | WMS: Paired Associations               |         |
| Duff and Hammers, 2022          | MCI: 93     | LO Mean (SD) of 1.5 (0.1) years | 2                 |                   | SDMT                                   |         |
|                                 |             |            |                    |                   | TMT-A                                  |         |
|                                 |             |            |                    |                   | TMT-B                                  |         |
|                                 |             |            |                    |                   | BVMT-R (total and delayed recall)      |         |
|                                 |             |            |                    |                   | HVLT (total and delayed recall)        |         |
| Duff et al. (2007)              | MCI: 8      | LO 2 weeks  | 2                 |                   | BVMT-R (total recall)                  | Lack of statistical testing. |
|                                 |             |            |                    |                   | MMSE                                   |         |
| Duff et al. (2011)              | MCI: 51     | LO 1 week   | 2                 |                   | SDMT                                   | Lack of statistical testing. |
|                                 | HC: 57      |            |                    |                   | TMT-A                                  |         |
|                                 |             |            |                    |                   | TMT-B                                  |         |
|                                 |             |            |                    |                   | BVMT-R (total and delayed recall)      |         |
|                                 |             |            |                    |                   | HVLT (total and delayed recall)        |         |
| Duff et al. (2012)              | Dementia, MCI: 61 | LO Single session | 2                 |                   | HVLT                                   | Lack of statistical testing. |
|                                 | AD: 61      |            |                    |                   | WAIS: Coding                           |         |
| Duff et al. (2014)              | MCI: 10     | LO 1 week   | 2                 |                   | BVMT-R (delayed recall)                | Lack of statistical testing. |
|                                 | HC: 15      |            |                    |                   | SDMT                                   |         |
|                                 |             |            |                    |                   | TMT-A                                  |         |
|                                 |             |            |                    |                   | TMT-B                                  |         |
|                                 |             |            |                    |                   | BVMT-R (total and delayed recall)      |         |
|                                 |             |            |                    |                   | HVLT (total and delayed recall)        |         |
| Duff et al. (2015)              | MCI: 10     | LO Approx. 1 week | 2                 |                   | SDMT                                   | Lack of statistical testing. |
|                                 | HC: 15      |            |                    |                   | TMT-A                                  |         |
|                                 |             |            |                    |                   | TMT-B                                  |         |
|                                 |             |            |                    |                   | BVMT-R (total and delayed recall)      |         |
|                                 |             |            |                    |                   | HVLT (total and delayed recall)        |         |

(continued on next page)
Table 7 (continued)

| Study                | Sample size | Study type | Follow-up duration | # test iterations | Practice effects in cohort of interest | Comment                                                                 |
|----------------------|-------------|------------|--------------------|-------------------|----------------------------------------|-------------------------------------------------------------------------|
| **Duff et al. (2017)** | MCI: 58     | LO         | 1 week             | 2                 | BVMT-R (total and delayed recall)     | • SDMT                                                                  |
|                      |             |            |                    |                   | HVL (total and delayed recall)        | • TMT-A                                                                 |
|                      |             | Add. cohort|                    |                   | Practice effects                        |                                                                         |
|                      |             |            |                    |                   | Initial effects                        |                                                                         |
|                      |             |            |                    |                   | Inconclusive                           |                                                                         |
|                      |             |            |                    |                   | No improvement                          |                                                                         |
| **Duff et al. (2018)** | MCI: 17     | LO         | Approx. 1 week     | 2                 | BVMT-R (total and delayed recall)     | • SDMT                                                                  |
|                      |             |            |                    |                   | HVL (total and delayed recall)        | • TMT-A                                                                 |
|                      |             | HC: 8      |                    |                   | Practice effects                        |                                                                         |
|                      |             | Add. cohort|                    |                   | Initial effects                        |                                                                         |
|                      |             |            |                    |                   | Inconclusive                           |                                                                         |
|                      |             |            |                    |                   | No improvement                          |                                                                         |
| **Elman et al. (2018)** | MCI and HC: 995 | LO | 6 years | 2 | Stroop-Color Test | Stroop-Word Test  | A trend towards improvement was observed on the Stroop Word Test, Digit Span (forwards condition only), Visual Reproduction Test (immediate recall only), and the Matrix Reasoning tests. |
|                      |             |            |                    |                   | CVLT (total and short delayed recall) | • WMS: Digit Span (forwards)                                          |
|                      |             |            |                    |                   | WMS: Digit Span (immediate recall)    | • WMS: Visual Reproduction Test (immediate recall)                       |
|                      |             |            |                    |                   | WMS: Spatial Span (total and backwards)| WASI: Matrix Reasoning                                                  |
|                      |             |            |                    |                   | WMS: Letter-Number Sequencing          |                                                                         |
|                      |             | Add. cohort|                    |                   | Practice effects                        |                                                                         |
|                      |             |            |                    |                   | Initial effects                        |                                                                         |
|                      |             |            |                    |                   | Inconclusive                           |                                                                         |
|                      |             |            |                    |                   | No improvement                          |                                                                         |
| **Frank et al. (1996)** | AD: 56     | LO         | Approx. 2.4 years  | 2                 | WMS: Visual Reproduction (immediate and delayed recall; at risk for AD only) | Practice effects on the Visual Reproduction test were only observed in patients with MCI, but not in patients at risk of developing AD. On the TMT-B, Visual Reproduction (delayed recall in AD patients) and MMSE, practice effects were only observed in specific sex subgroups (male vs female). |
|                      |             | HC: 242    |                    |                   | TMT-B                                   |                                                                         |
|                      |             | Add. cohort|                    |                   | WMS: Visual Reproduction (immediate recall; AD only) |                                                                         |
|                      |             |            |                    |                   | Practice effects                        |                                                                         |
|                      |             |            |                    |                   | Initial effects                        |                                                                         |
|                      |             |            |                    |                   | Inconclusive                           |                                                                         |
|                      |             |            |                    |                   | No improvement                          |                                                                         |
| **Gavett et al. (2016)** | MCI: 72    | LO         | 5 years            | 5                 | WMS: Logical Memory (immediate and delayed recall) | Practice effects were only observed in patients with MCI, but not in patients with AD. |
|                      |             | HC: 96     |                    |                   | TMT-B                                   |                                                                         |
|                      |             | Add. cohort|                    |                   | Practice effects                        |                                                                         |
|                      |             |            |                    |                   | Initial effects                        |                                                                         |
|                      |             |            |                    |                   | Inconclusive                           |                                                                         |
|                      |             |            |                    |                   | No improvement                          |                                                                         |
| **Gross et al. (2018)** | AD: 990    | LO         | 2.4 years          | ≤7                | MMSE                                    | A random effects model analysis revealed an overall main retest (practice) effect over time. |
|                      |             |            |                    |                   | Practice effects                        |                                                                         |
|                      |             | Add. cohort|                    |                   | Initial effects                        |                                                                         |
|                      |             |            |                    |                   | Inconclusive                           |                                                                         |
|                      |             |            |                    |                   | No improvement                          |                                                                         |
Table 7 (continued)

| Study                                      | Sample size | Study type | Follow-up duration | # test iterations | Practice effects in cohort of interest | Comment                                                                                           |
|--------------------------------------------|-------------|------------|--------------------|-------------------|---------------------------------------|--------------------------------------------------------------------------------------------------|
| Hammers et al. (2011)                      | MCI: 20     | HC: 23     | LO                 | Single session    | 2                                     | Continuous effects: CogState: OBK accuracy (all cohorts)                                       |
|                                            | AD: 52      |            |                    |                   |                                       | Practice effects on the OBK reaction time task were observed only in patients with MCI or DLB. |
|                                            | Dementia (incl. DLB, FTD): 19 |          |                    |                   |                                       | CogState: Identification                                                                          |
|                                            |              |            |                    |                   |                                       | CogState: One Card Learning                                                                       |
|                                            |              |            |                    |                   |                                       | CogState: Associative Learning                                                                     |
| Rosas et al. (2020)                        | MCI: 270d    | HC: 46e    | LO                 | Mean (SD) of 25.96 (11.28) months | 2                                     | Continuous effects: TMT-A                                                                         |
|                                            | SCD: 42d     |            |                    |                   |                                       | Practice effects were indirectly inferred from reliable change index analyses.                     |
|                                            |              |            |                    |                   |                                       | TMT-B, Stoop Word Test                                                                           |
|                                            |              |            |                    |                   |                                       | Stoop Color Test                                                                                 |
|                                            |              |            |                    |                   |                                       | WAIS: Digit Symbol                                                                               |
| Teasdale et al. (2016)                     | MCI: 15      |            | LO                 | 6 months          | 6                                     | Continuous effects: Driving simulator                                                            |
|                                            |              |            |                    |                   |                                       | Practice effects observed only during training phase during which feedback was given.            |

AD, Alzheimer’s disease; Approx., approximately; BVMT-R, Brief Visuospatial Memory Test-Revised; CVLT, California Verbal Learning Test; DLB, dementia with Lewis Bodies; FTD, frontotemporal dementia; HC, healthy controls; HVLT, Hopkins Verbal Learning Test; IDM, divided attention task; MCI, mild cognitive impairment; MMSE, Mini-Mental State Examination; OBK, One-Back Test; SCD, subjective cognitive decline; SDMT, Symbol Digit Modalities Test; SMC, subjective memory complaint; TMT, Trail-Making Test; WASI, Wechsler Abbreviated Scale of Intelligence; WMS, Wechsler Memory Scale.

a ‘Continuous effects’ is defined as a continuous improvement in test performance for ≥4 test administrations, with test performance continuing to improve up to the last test administered. By definition this can only be applied to studies that administered the test at least 4 times. In all other instances, practice effects are described as either ‘initial effect’ if clear signs of practice effects were evident; ‘inconclusive’ if practice effects were observed for a selection of test metrics, in a subgroup of patients only, or if other reasons may explain the improvement in test performance (contribution of other tests in composite scores, association with additional training or treatment etc.); or ‘no effect’ if no improvement in test performance was observed.

b Based on clinical rating at the end of the study.

c Data from 1,220 and 995 patients were available for visit 1 and 2, of which 11.0% and 15.2% (after correcting for practice effects) were diagnosed with mild cognitive impairment, respectively.

d At follow-up, 48 participants were diagnosed with Alzheimer’s disease, 200 with mild cognitive impairment and 64 with subjective cognitive decline, while 46 participants were considered as cognitively healthy.
3.7.5. Run-in period

Since improvements due to practice are typically strongest between the first few test iterations, a run-in or familiarization period prior to taking the baseline assessments has been suggested to reduce the magnitude of practice effects (Beglinger et al., 2014a; Beglinger et al., 2014b; Cohen et al., 2000; Stout et al., 2014; Sormani et al., 2019). Such a run-in period would allow patients to become fully familiar with the test and test conditions and to reach steady-state performance prior to their baseline assessment, thereby preventing post-baseline practice effects (Patzold et al., 2002). For the success of a run-in period, it is therefore critical to administer a sufficient number of tests prior to the baseline assessment, thereby preventing post-baseline practice effects (Patzold et al., 2002). For the success of a run-in period, it is therefore critical to administer a sufficient number of tests prior to the baseline assessment. However, many studies included in this analysis have administered only two or three test iterations (Tables 4, 5, 6, and 7). Such a run-in period would allow patients to become fully familiar with the test and test conditions and to reach steady-state performance prior to their baseline assessment, thereby preventing post-baseline practice effects (Patzold et al., 2002). For the success of a run-in period, it is therefore critical to administer a sufficient number of tests prior to the baseline assessment. However, many studies included in this analysis have administered only two or three test iterations (Tables 4, 5, 6, and 7). This makes it more challenging to establish the minimum number of tests required for the run-in period for each of the four disease areas. For instance, in patients with multiple sclerosis, two to three pre-baseline assessments have been recommended for the PASAT (Cohen et al., 2000; Rosti-Ottajärvi et al., 2008). This may not be sufficient considering that a trend of continuous improvement beyond the third test iteration was observed in Glanz et al. (2012). Similarly, Gavett et al. (2016) argued that the previously recommended 2 or 3 pre-baseline assessments with the Wechsler Memory Scale may not be sufficient to prevent further practice effects as patients with mild cognitive impairment showed continuous improvements over all 5 test iterations (Gavett et al., 2016). In addition, the inter-test interval during the run-in period can also impact the likelihood of post-baseline practice effects (Beglinger et al., 2014a). Finally, implementing a run-in period will increase the burden of the patient and increase the cost and time needed to run clinical trials.

4. Discussion and outlook

Practice effects are a common phenomenon associated with the repeated administration of performance outcome measures (Tables 4, 5, 6, and 7). Despite the research conducted on practice effects, some gaps still remain:

- Many different approaches to identify, assess and study practice effects have been applied, which complicates a comparison across studies.
- Most studies defined practice effects on the basis of improved test performance at retest. This does not allow for practice effects and functional decline (and other longitudinal effects), which is commonly observed in patients with chronic neurologic disorders, to coincide. Such changes in functional ability limit the ability to detect and to account for practice effects. Thus, optimal methods to distinguish between practice effects and functional decline, but also changes in motivation and fatigue, functional recovery, and treatment effects need to be further investigated.
- The possible impact of previous exposure on the ability to detect further practice effects was largely unaddressed.
• The temporal dynamics of practice effects has not been studied in detail and further research could expand our understanding how practice effects vary over time.
• Practice effects on an individual patient level have not been fully characterized.
• The clinically meaningful information contained within practice effects remains unclear and further research is needed to establish their usefulness in guiding disease management.
• Finally, further research into the possible impact of the more granular datasets collected with digital performance outcome measures on practice effects is needed.

4.1. Comparing practice effects across studies

This review revealed that practice effects were consistently observed across studies for measures of information processing speed or upper extremity function. In contrast, results were more mixed for other measures. Many different factors can contribute towards these mixed results. One possible explanation lies in the nature of the test. Changing the sequence of items by using an alternate form, if available, can reduce the magnitude of practice effects that are driven by item learning (Benedict, 2005). Conversely, the use of the same form increases the magnitude of practice effects (Duff et al., 2018). Differences in the patient characteristics may have also contributed to the mixed results. Studies in patients with mild cognitive impairment, Alzheimer's disease or other forms of dementia suggest that more severe disease is associated with less consistent or weaker practice effects (Campos-Magdaleno et al., 2017; Duff et al., 2012; Frank et al., 1996; Gavett et al., 2016; Hammers et al., 2011). However, Rosas et al., (2020) showed that the proportion of patients showing practice effects increases with disease severity. This points towards some unobserved confounders explaining the differences between studies. Such confounders could include cognitive training or therapeutic interventions between the assessments (Rosas et al., 2020). Others have suggested that a poor baseline performance lends itself to larger margins for improvement, and therefore, to stronger practice effects (Rabbit et al., 2004). But also any previous exposure to the test could impact the ability to detect and quantify practice effects (see also section ‘4.3 Impact of previous exposure on the ability to detect practice effects’). Finally, the different approaches used to identify, assess and study practice effects (Table S1) further complicate a direct comparison across studies and may partially explain why practice effects were observed in some but not in all studies. Such a comparison would have benefited from a more standardized approach.

4.2. Distinguishing practice effects from other effects

A distinction between practice effects and longitudinal effects such as changes in motivation and fatigue, functional recovery and treatment effects was not always possible (Giedraitiene and Kaubrys, 2019; Reilly and Hynes, 2018; Vogt et al., 2009). While changes in motivation and fatigue might coincide with practice effects with both short and long inter-test intervals, functional changes and treatment effects may increasingly impact the ability to discern practice effects the longer the inter-test intervals are. Consequently, stable performance does not necessarily guarantee functional stability as practice effects may mask true functional decline (Elman et al., 2018). Assessing practice effects in the non-interventional cohort separately from the interventional cohort can help to disentangle practice effects from treatment or other interventional effects (Buell et al., 2015; Patzold et al., 2002; Turner et al., 2016; Vogt et al., 2009).

4.3. Impact of previous exposure on the ability to detect practice effects

Most studies did not specify the previous exposure to the investigated performance outcome measures. Only five studies stipulated requirements with regard to previous exposure, or lack thereof, in their inclusion and exclusion criteria (Cohen et al., 2000; Elman et al., 2018; Fuchs et al., 2020; Nagels et al., 2008; Patzold et al., 2002). Two additional studies included a run-in period or practice items prior to the baseline assessment to minimize potential practice effects (Patzold et al., 2002; Snowden et al., 2001). Thus, it is feasible that previous exposure impacted the ability to detect further practice effects, which may partially explain the mixed results observed on some of the performance outcome measures.

4.4. Temporal dynamics of practice effects

Another limitation is that practice effects were often assumed to be constant, or linear, over time. However, the temporal dynamics of practice effects, including the duration until steady-state performance is reached and the optimal inter-test interval to minimize the impact of practice effects, has not been studied in detail. Only two studies quantified the duration of the practice phase (Pham et al., 2021; Prince et al., 2018). In addition, it has been shown that practice gains are not linear over time: the gain from practice is greatest over the first few test repetitions and gradually becomes smaller as the number of test repetitions increases (Glanz et al., 2012). In other words, overall practice effects can vary as a function of time or number of test iterations. Strategies that explicitly take this non-linear nature of practice gains into account could help to further improve the accuracy of the interpretation of longitudinal datasets.

The non-linear nature of practice effects resides in the rapid gain in performance observed in the first few test iterations, which cannot be linearly modelled with the later stabilization of test performance as the number of iteration increases. This dynamic has been highlighted for example by Pham et al. (2021). However, methodologies explicitly characterizing the non-linear nature of practice are still few and far between. Yet, as this manuscript is a review of the current state of the literature, it wouldn't be appropriate to propose such methodologies without also presenting results illustrating and characterizing such novel approaches.

4.5. Practice effects on a group versus patient level

Similarly, most studies limited their assessment of practice effects to a group-level analysis. Questions such as ‘is the improvement in test performance statistically significant?’ or ‘is the change of the cohort greater (or smaller) than expected based on the change seen in a normative population?’, however, do not take differences in practice effects between individual patients into account. Only few studies acknowledged that practice effects may differ from patient to patient and performed their analyses in subgroups of patients defined by their practice effects response or on an individual patient level (Duff et al. 2014, 2017; Pham et al., 2021; Prince et al., 2018; Rosas et al., 2020; Somarni et al., 2019; Turner et al., 2016; Utz et al., 2016). Analyses of longitudinal data in daily clinical practice could however benefit from methods to study practice effects on an individual patient level. This will require a more granular dataset than typically obtained with traditional clinician-administered, in-clinic performance outcome measures. As discussed further below in section ‘4.8 Outlook’, remotely administered digital, sensor-based performance outcome measures could help to collect sufficient data to study practice effects on an individual patient level.

4.6. Clinical impact of practice effects

Some have argued that practice effects contain clinically meaningful information and should not be regarded as only a source of unwanted variance. For example, cognitive impairment may be expressed as a diminished capacity to learn, resulting in weaker practice effects (Gavett et al., 2016). Short inter-test intervals, in particular, have been purposely used to elicit practice effects that capture clinically relevant information
treatment outcomes in multiple sclerosis (Sormani et al., 2019). More-
et al., 2014). Weaker practice effects have been associated with worse
prognosis in mild cognitive impairment (Duff et al., 2011) and worse
effects elicited with inter-test intervals as short as one week are more
misdiagnosis or misclassi
suf
scales, or performance outcome measures, associated with minimal
timely manner (Rae-Grant et al., 2015). The use of sensitive disability
mended to optimally track the disease and detect disease progression in a
clinical management of multiple sclerosis. Nonetheless, regular assess-
4.7. Mitigating practice effects

Despite the research effort, no consensus has been reached on an
effectiveness of an intervention is assessed with performance
outcome measures (see for example Pliskin et al. (1996)). Thus, it is
important for clinical trials to have effective measures in place that ac-
count for and mitigate practice effects.

4.8. Outlook

Digital, sensor-based tests are increasingly being studied for the
assessment of functional ability (Pham et al., 2021; Prince et al., 2018).
In contrast to traditional, clinician-administered tests, digital tests are
typically self-administered at home and enable short inter-test intervals
with prolonged study durations (Prince et al., 2018; Westin et al., 2010).
This is both a curse and a blessing as it may expose potential practice
effects more prominently while offering a more granular and ecologically
valid assessment of functional ability through time. The increased
granularity also allows a more objective comparison of absolute test
performance across and within subjects as it can be assumed that the
impact of practice effects is negligible once a certain number of test it-
erations has been reached (Cook et al., 2004).

In order to leverage this advantage of digital tests and establish
suitable baseline performance in non-digital tests, an in-depth analysis of
the effect of practice and gold-standard statistical methods is required.
First, different metrics and approaches to study practice effects will need
to be investigated to establish the metrics that describe practice effects
optimally, in particular in the emerging field of digital tests. This includes
assessing the applicability of learning curve models to characterize
practice effects to study the temporal dynamics of practice effects. Such a
model has been previously applied to data obtained from smartphone-
based SDMT (Pham et al., 2021). Second, it is yet to be shown whether
the higher granularity of digital tests results in a better separation of
practice effects and other effects impacting test performance such as
functional changes or treatment effects. Third, it has been previously
suggested that practice effects may contain clinically relevant informa-
tion (Duff et al., 2011, 2014, 2018; Sormani et al., 2019). Future work
will therefore need to investigate whether digital tests can be leveraged
to extract such information. Finally, digital tests could also be used to
disentangle test features and their underlying functions, for example,
sensorimotor, cognitive and memorization processes that show practice
effects from those that do not. This can be leveraged to develop future
tests and test features that are resistant to practice effects, thereby
simplifying the assessment of a subject’s functional capacity. These ef-
forts will provide us with a more detailed understanding of practice ef-
effects, allows us to more accurately interpret longitudinal data and
possibly help us to separate the unwanted noise introduced by practice
effects from the clinical meaningful information contained within them.

4.9. Limitations

Practice effects were only considered if the improved test perfor-
manecer, or test performances that were better than expected, due to
practice or the repetition of a task could not be explained by means,
including interventional effects, functional recovery, or changes in
motivation and fatigue levels. This definition limited the ability to
identify practice effects that occurred within the same time frame as
functional decline (or other longitudinal effects that result in worsened
test performance), as can be expected in patients with chronic neurologic
disorders. However, this limitation is not only a limitation of this review,
but also a general limitation of the study of practice effects. Furthermore,
most studies enrolled predominantly white, highly educated participants,
which limits the generalizability of the findings.

5. Conclusions

The variance introduced by practice effects is shared by many per-
formance outcome measures and could ultimately be addressed by the
thorough characterization and evaluation of such alterations for specific
subject populations, study designs, test activities, and delivery proced-
ures. Due to its prevalence, an analysis of the presence and magnitude of
practice effects on inter-individual and intra-individual data obtained
from repeated assessments should be expected. Additionally, mitigation
strategies should be in place from study design to data analysis, especially
if practice effects interact with the studied intervention. The failure to do so may result in misdiagnosis or inaccurate interpretation of clinical data. In light of the recent development of digital, sensor-based test batteries and their associated higher number of test iterations, there is renewed need for strategies to assess practice effects and mitigate their impact on the interpretation of clinical data. In particular, the much higher granularity made possible with digital tests offers a new opportunity to properly characterize and tackle the impact of practice effects on performance outcome measures, including deconvolving practice effects from true functional changes and treatment effects in clinical trials and clinical practice. Future work should, therefore, aim to identify optimal metrics for detecting and characterizing practice effects and their properties in such highly granular datasets. In addition, the analysis of practice effects should also be leveraged to guide adequate study design, data analysis strategy, and the selection of novel digital test features that are resistant to practice effects.

Declarations

Author contribution statement

All authors listed have significantly contributed to the development and the writing of this article.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability statement

Data included in article/supp. material/referenced in article.

Declaration of interest's statement

Sven P. Holm is a contractor for F. Hoffmann-La Roche Ltd. Arndna M. Woller is an employee of F. Hoffmann-La Roche Ltd. Grégoire H.S. Pointeau is an employee and shareholder of F. Hoffmann-La Roche Ltd. Florian Lipsmeier is an employee of F. Hoffmann-La Roche Ltd. Michael Lindemann is a consultant for F. Hoffmann-La Roche Ltd. via Inovigate.

Additional information

Supplementary content related to this article has been published online at https://doi.org/10.1016/j.helyon.2022.e10259.

References

Arvantiakiz, Z., Shah, R.C., Bennett, D.A., 2019. Diagnosis and management of dementia: review. JAMA 322, 1589–1599.

Bachoud-Levi, A.C., Maione, P., Bartolomeo, P., Boizet, M.F., Dalla Barba, G., Ergis, A.M., Badie, S., Degas, J.D., Ceano, P., Peschanski, M., 2001. Retest effects and cognitive decline in longitudinal follow-up of patients with early HD. Neurology 56, 1052–1058.

Barker-Collo, S.L., 2005. Within session practice effects on the PASAT in clients with cognitive impairment: practice effects and meaningful changes. Front. Psychol. 8, 1231.

Clark, J.J., Mohr, E., Chase, T.N., 1991. Clinical trials in dementia: learning effects with repeated testing. J. Psychiatry Neurosci. 16, 1–4.

Cohen, J., 1992. A power primer. Psychol. Bull. 112, 155–159.

Cohen, J.A., Cutter, G.R., Fischer, J.S., Goodman, A.D., Heidenreich, F.R., Jak, A.J., Knier, J.E., Kooijman, M.F., Lull, J.M., Sandrock, A.W., Simon, J.H., Simonian, N.A., Whitaker, J.N., 2001. Use of the multiple sclerosis functional composite as an outcome measure in a phase 3 clinical trial. Arch. Neurol. 58, 961–967.

Cohen, J.A., Fischer, J.S., Bolbruch, D.M., Jak, A.J., Knier, J.E., Mertz, L.A., Skaramagas, T.T., Cutter, G.R., 2000. Intrarater and interrater reliability of the MSFC functional composite outcome measure. Neurology 54, 802–806.

Cook, J.A., Ramsay, C.R., Fayers, P., 2004. Statistical evaluation of learning curve effects in surgical trials. Clin. Trials 1, 421–427.

Dela, D.C., Kramer, J.H., Kaplan, E., Ober, B.A., 1987. The California Verbal Learning Test, Research Edition. Adult Version. The Psychological Corporation, San Antonio, TX.

Duff, K., Anderson, J.S., Mallik, A.K., Suhrie, K.R., Atkinson, T.J., Dalley, R.C.A., Morimoto, S.S., Hoffman, J.M., 2018. Short-term repeat cognitive testing and its relationship to hippocampal volumes in older adults. J. Clin. Neurosci. 57, 121–125.

Elman, J.A., Jak, A.J., Panizzon, M.S., Tu, X.M., Chen, T., Reynolds, C.A., Gustavson, D.E., Duff, K., Chelune, G., Dennett, K., 2012. Within-session practice effects in patients with MCI. J. Clin. Exp. Neuropsychol. 39, 396–407.

Elman, J.A., Jak, A.J., Suhrie, K.R., Dalley, B.C., Schafer, S.Y., Hammers, D.B., 2017. Short-term practice effects in mild cognitive impairment: evaluating different methods of change. J. Clin. Exp. Neuropsychol. 39, 396–407.

Duff, K., Keshavarz, S., Osler, D.M., McCaffrey, R.J., Haase, R.F., Westervelt, H.J., Langbehn, D.R., Paulson, J.S., Hsu, G., 2007. Practice effects in the prediction of long-term cognitive outcome in three patient samples: a novel prognostic index. Arch. Clin. Neuropsychol. 22, 15–24.

Duff, K., Chehune, G., Bennet, D.K., 2012. Within-session practice effects in patients referred for suspected dementia. Dement. Geriatr. Cognit. Disord. 33, 245–249.

Duff, K., Foster, N.L., Hoffman, J.M., 2014. Practice effects and amyloid deposition: preliminary data on a method for enriching samples in clinical trials. Alzheimer Dis. Assoc. Disord. 28, 247–252.

Duff, K., Hammers, D.B., 2022. Practice effects in mild cognitive impairment: a validation of Galalma et al. (2012). Clin. Neuropsychol. 36 (3), 571–583.

Duff, K., Horn, K.P., Foster, N.L., Hoffman, J.M., 2015. Short-term practice effects and brain hypometabolism: preliminary data from an FDG PET study. Arch. Clin. Neuropsychol. 30, 264–270.

Duff, K., Lyketsos, C.G., Beglinger, L.J., Chehune, G., Moser, D.J., Arndt, S., Schultz, S.K., Paulson, J.S., Petersen, R.C., McCaffrey, R.J., 2011. Practice effects predict cognitive outcome in amnestic mild cognitive impairment. Am. J. Geriatr. Psychi. 19, 932–939.

Elman, J.A., Jak, A.J., Minzoni, M.S., Tu, X.M., Chen, T., Reynolds, C.A., Gustavson, D.E., Duff, K., Hatton, S.N., Toomey, S.H., Migliorini, R., Xian, H., Lyons, M.J., Kremen, W.S., 2018. Underdiagnosis of mild cognitive impairment: a consequence of ignoring practice effects. Alzheimers Dement (Amst) 10, 372–381.

Elwood, R.W., 1995. The California Verbal Learning Test: psychometric characteristics and clinical application. Neuropsychol. Rev. 5, 173–201.

Elmer, D.M., Kautzki, T., Caruso, L.S., Benedict, R.H., Foley, F.W., Wilken, J., Cadavid, D., Deluca, J., 2014. Reliability of a cognitive endpoint for use in a multiple sclerosis pharmacoeconomic trial. J. Neurol. Sci. 340, 123–129.

Edahahi, A., Riyahi-Alam, S., Roostaei, T., Haeri, G., Aghaei, A., Aidi, M.R., Pourtemad, H.R., Zarei, M., Farhang, S., Saiedi, R., Nazeri, A., Gajjehi, H., Etesam, F., Falleti, M.G., Maruff, P., Collie, A., Darby, D.G., 2006. Practice effects associated with the repeated assessment of cognitive function using the Cogstate battery at 10-minute, 1-week and one month test-retest intervals. J. Clin. Exp. Neuropsychol. 28, 1095–1102.
Glazn, B.I., Healy, B.C., Hvid, L.E., Chitnis, T., Weiner, H.L., 2012. Cognitive deterioration in patients with early multiple sclerosis: a 5-year study. J. Neurology. Psychiatry 83, 38–42.

Goldberg, T.E., Harvey, P.D., Wesnes, K.A., Snyder, P.J., Schneider, L.S., 2015. Practice effects on cognitive functions in clinical contexts. Clin. Neuropsychol. 24, 1267–1278.

Higgins, C.J., Wheelock, V.L., Levine, D., King, D.S., Pappas, C.T., Sigvardt, K.A., 2010. Official position of the American Academy of Clinical Neuropsychology on serial neuropsychological assessments: the utility and challenges of repeat test administration in clinical neuropsychology. Clin. Neuropsychol. 24, 103–111.

Gronwall, D.M., 1977. Paced auditory serial-addition task: a measure of recovery from concussion. Percept. Mot. Skills 44, 367–373.

Gross, A.L., Chu, N., Anderson, L., Glymour, M.M., Jones, R.N., Diseases, C.A.M., 2018. Do catastrophic events shape neurocognitive trajectories? A study of a 17-year longitudinal study of cognitive aging. J. Gerontol. B Psychol. Sci. Soc. Sci. 73, 7–15.

Prince, J., Arora, S., de Vos, M., 2018. Big data in Parkinson’s disease: using smartphones to remotely detect longitudinal disease progression. Physiol. Meas. 39, 044005.

Rae-Grant, A., Bennett, A., Sanders, A.E., Phipps, M., Cheng, E., Bever, C., 2015. Quality improvement in neurology: multiple sclerosis quality measures: executive summary. Am. J. Phys Med. Rehabil. 94, 1053–1065.

Rao, S.M., Galonio, S., Sokolskis, M., McGinley, M., Freiberger, J., Weber, M., Dey, T., Mourany, L., Schindler, D., Reecie, C., Miller, D.M., Bethoux, F., Berel, R.A., Williams, J.R., Levit, N., Phillips, G.A., Rhodes, J.K., Alberts, J., Rudic, R.A., 2020. Multiple sclerosis performance and validation of self-administered neuropsychometric modules. Eur. J. Neurolog.

Rao, S.M., Leo, G.J., Haughton, V.M., St Aubin-Faubert, P., Bernardin, L., 1989. Correlation of magnetic resonance imaging with neuropsychological testing in multiple sclerosis. Neuropsychologia 27, 371–390.

Reilly, S., Hynes, S.M., 2018. A cognitive occupation-based programme for people with multiple sclerosis: a study to test feasibility and clinical outcomes. Occup. Ther. Int. 25, 59–66.

Ritti-Ebere, N., Aiyer, E., Wood, F., 2015. A distinctive pattern of cognitive disorders during multiple sclerosis relapse and recovery based on computerized CANTAB tests. Front. Neurol. 5, 107.

RNA, A.J., Hardy, J., Revesz, T., 2009. Parkinson’s disease. Lancet 373, 2055.

Salthouse, T.A., 2011. What cognitive abilities are involved in trail-making performance? Intelligence 39, 222–232.

Sarno, P., Holm et al. Heliyon 8 (2022) e10259

Schoebel, L.E., Seel, M.M., Hesse, R.J., Stern, C.E., 2003. An FMRI study of the role of the medial temporal lobe in implicit and explicit sequence learning. Neurology 37, 1013–1025.

Schindler, D., Reecie, C., Miller, D.M., Bethoux, F., Berel, R.A., Williams, J.R., Levit, N., Phillips, G.A., Rhodes, J.K., Alberts, J., Rudic, R.A., 2020. Multiple sclerosis performance and validation of self-administered neuropsychometric modules. Eur. J. Neurolog.

Shuttleworth, R., Mathews, J., 1994. The reliability and validity of the TMT-A and TMT-B in a cohort of healthy participants. Acta Neurolog. Scand. 90, 335–340.

Silverman, E.M., Hersh, A.M., Bollinger, J., 2007. The sensitivity and specificity of clinical diagnosis of Parkinson disease. Neurology 68, 1553–1558.

Smith, A., 1982. Symbol Digit Modalities Test: Manual. Western Psychological Services, Los Angeles, CA, USA.

Sneddon, J., Graudziutis, T., Roberts, C., 2010. Huntington’s disease: a clinical review. Orphanet J. Rare Dis. 5, 40.

Spreng, R.N., Mar, R.A., Kim, K., 2012. A network-based perspective on thequantity and quality of functional connectivity in human brain resting state networks. Front. Hum. Neurosci. 6, 18.

Spencer, J.T., 2006. Psychopharmacology: the third generation of progress. Science 314, 1596–1601.

Steiner, E., 1961. Purdue Pegboard Examiner Manual. Science Research Associates, Chicago, IL.

Stoodley, K.S., Lee, D.H., Linker, R.A., €ockinglin, M., Gschwind, L., Opwis, K., Penner, I.K., €arkinglin, M., Gschwind, L., Opwis, K., Penner, I.K., 2010. Repeated assessment of key clinical walking measures can induce confounding effects in patients with multiple sclerosis relapse and recovery based on computerized CANTAB tests. Front. Neurol. 1, 103.

Stern, J.E., 1976. Conceptual and measurement issues in the assessment of Alzheimer’s disease. J. Clin. Exp. Neuropsychol. 8, 471–500.

Stout, J.C., Queller, S., Baker, K.N., Cowlishaw, S., Sampaio, C., Fitzer-Attas, C., Borovsky, B., Investigators, I.L.-C., 2014. HD-CAB: a cognitive assessment battery for clinical trials in Huntington’s disease. Mov. Disord. 29, 1281–1288.

Sperling, R.A., 2010. Huntington’s disease: a clinical review. Orphanet J. Rare Dis. 5, 40.

Talairach, J., 1988. Co-Planar Stereotaxic Atlas of the Human Brain. Thieme Medical Publishers, New York.

Turner, T.H., Resnoff, J.B., Elm, J., Duppstadt-Delambo, A., Hinson, V.K., 2016. Voluntaryness of reliable change indices: a test of variability in Parkinson's disease with mild cognitive impairment. Appl. Neuropsychol. Adult 23, 399–410.

Utz, K.S., Hanks, T.M., Jung, L., Lämmer, A., Waschbisch, A., Lee, D.H., Linker, R.A., Schenk, T., 2013. Visual search as a tool for a quick and reliable assessment of cognitive functions in patients with multiple sclerosis. Front. Hum. Neurosci. 7, 635.

Vogt, A., Kopp, L., Labare, P., Stocklin, M., Gschwind, L., Opwis, K., Penner, L.K., 2009. Working memory training in patients with multiple sclerosis - comparison of two different training schedules. Restor. Neurol. Neurosci. 27, 225–235.

Wechsler, D., 2009. Wechsler Memory Scale. Pearson, San Antonio, TX.

Wechsler, D., 2008. Wechsler Adult Intelligence Scale. Pearson, San Antonio, TX.

Wechsler, D., 2009. Wechsler Memory Scale. Pearson, San Antonio, TX.

Wechsler, D., 2008. Wechsler Adult Intelligence Scale. Pearson, San Antonio, TX.

Wechsler, D., 2008. Wechsler Adult Intelligence Scale. Pearson, San Antonio, TX.

Wechsler, D., 2008. Wechsler Adult Intelligence Scale. Pearson, San Antonio, TX.