medRxiv preprint doi: https://doi.org/10.1101/2021.08.28.21262764; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

# Melodic Intonation Therapy for aphasia: A multi-level meta-analysis

Dr. Tudor Popescu[1,2*], Prof. Benjamin Stahl[3-6], Prof. Brenton M. Wiernik[7], Ms Hannah Helm[3], Ms Michaela Zemanek[1], Mr Felix Haiduk[1], Mr Theresa Matzinger[1,8], Prof. Roland Beisteiner[2], Prof. Tecumseh W. Fitch[1]

[1] University of Vienna, Department of Behavioural and Cognitive Biology, Vienna, Austria

[2] Medical University of Vienna, Department of Neurology, Vienna, Austria

[3] Medical School Berlin, Department of Psychology, Berlin, Germany

[4] Charité Universitätsmedizin, Department of Neurology, Berlin, Germany

[5] Max Planck Institute for Human Cognitive and Brain Sciences, Department of Neurophysics, Leipzig, Germany

[6] Universitätsmedizin Greifswald, Department of Neurology, Greifswald, Germany

[7] University of South Florida, Department of Psychology, Tampa, Florida, USA

[8] University of Vienna, Department of English, Vienna, Austria

[*] corresponding author: tudor.popescu@univie.ac.at; +43-1-4277-76101; Althanstrasse 14, 1090 Vienna, Austria

Manuscript word count (not including title, abstract, acknowledgment, references, tables, and figure legends): **2,915**



**NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.**

medRxiv preprint doi: https://doi.org/10.1101/2021.08.28.21262764; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

## KEY POINTS (MAX 100 WORDS)

QUESTION: What determines the efficacy of Melodic Intonation Therapy (MIT), arguably the best-known treatment programme for individuals with neurological communication disorders?

FINDINGS: MIT's effect size was modulated by the psychometric quality of outcomes, use of randomisation and control groups, and the number of months post-stroke at the time of testing. Language performance improved significantly on trained items, less for everyday communication ability on untrained items.

MEANING: Our findings emphasise the importance of appropriate outcomes and rigorous study design to obtain realistic effect size estimates. While MIT promotes performance on trained items, it appears to have limited impact on everyday communication ability.

## ABSTRACT (MAX 350 WORDS)

IMPORTANCE: Melodic Intonation Therapy (MIT) is a prominent rehabilitation programme for individuals with post-stroke aphasia. Despite substantial progress in recent years, the efficacy of MIT remains not fully understood.

OBJECTIVE: Based on *a-priori* hypotheses, the present meta-analysis investigated the efficacy of MIT while considering quality of outcomes (psychometrically validated *versus* unvalidated measures), experimental design (presence *versus* absence of randomisation and control group), influence of spontaneous recovery (quantified as number of months post-stroke), MIT version applied (original *versus* modified protocol), and level of generalisation (performance on trained *versus* untrained items).

DATA SOURCES: An extensive literature search in all major online databases, trials registers and the grey literature identified 606 studies (years searched: 1973–2021).

STUDY SELECTION: Inclusion criteria: randomised controlled trial (RCT) data or case reports on adults with aphasia; pre-post assessment of language performance. Exclusion criteria: substantial variation from

medRxiv preprint doi: https://doi.org/10.1101/2021.08.28.21262764; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

original MIT protocol; unvalidated outcomes, unless both trained and untrained items were compared; essential information not indicated/retrievable. Final sample: 22 studies.

DATA EXTRACTION AND SYNTHESIS: Following PRISMA guidelines, studies were double-coded. Multi-level mixed- and random-effects models were used to separately meta-analyse RCT and non-RCT data.

MAIN OUTCOMES AND MEASURES: Measures of language performance focused on aphasia severity, everyday communication ability, domain-general function, language comprehension, non-communicative language expression, and speech-motor planning.

RESULTS: Unvalidated outcomes appeared to attenuate MIT's effect size by a factor of 0.29–0.43 across study designs when compared to validated outcomes. Moreover, MIT's effect size was 5.7 times larger for non-RCT data compared to RCT data. Effect size also decreased with number of months post-stroke, suggesting confound through spontaneous recovery primarily within the first year post-stroke. In contrast, variation of the original MIT protocol did not systematically alter benefit from treatment. Crucially, analyses demonstrated significantly improved performance on trained and untrained items. The latter finding arose mainly from gains in repetition tasks, rather than other domains of verbal expression including everyday communication ability.

CONCLUSIONS AND RELEVANCE: Accounting for various methodological aspects, the current results confirm the promising role of MIT in improving language performance on trained items and in repetition tasks, while highlighting possible limitations in promoting everyday communication ability.

medRxiv preprint doi: https://doi.org/10.1101/2021.08.28.21262764; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

# 1. Introduction

Stroke survivors often experience a profound loss of communication skills, among them a syndrome known as aphasia. This syndrome may manifest as severe difficulty in verbal expression, referred to as 'non-fluent aphasia.' In addition, stroke survivors frequently suffer from impaired speech-motor planning. Known as 'apraxia of speech,' this syndrome typically occurs in combination with aphasia. Although about a third of individuals with neurological communication disorders do not recover completely[1], rehabilitation programmes can improve language performance even in the chronic stage of symptoms[2].

Melodic Intonation Therapy (MIT) is a prominent rehabilitation programme originally developed for individuals with non-fluent aphasia[3]. Drawing on the observation that individuals with neurological communication disorders are often able to sing entire pieces of text fluently[4–6], MIT uses melody, rhythm, vocal expression (in unison and alone), left-hand tapping, formulaic and non-formulaic verbal utterances, as well as other therapeutic elements, in a hierarchically structured protocol[7]. To date, randomised controlled trial (RCT) data have confirmed the efficacy of MIT on validated outcomes in the *late subacute* or *consolidation* stage of aphasia (i.e., up to 12 months after stroke)[8], but not in the *chronic* stage of aphasia (i.e., more than 6–12 months after stroke)[9].

From a methodological point of view, influences of spontaneous recovery are generally lower in the *chronic* stage of aphasia, as suggested by RCT data[10] and meta-analyses[11]. Therefore, it is important to consider stage of symptoms post-stroke. Moreover, speech-language therapy seeks to promote performance on untrained items. Consistent with this goal, the present work distinguishes progress on trained items—learning resulting from using the same set of utterances both during treatment and subsequent assessment—from the more desirable goal of attaining generalisation to untrained items, ideally in the context of everyday communication to ensure ecological validity[e.g., 12].

So far, there are several systematic reviews on MIT[e.g., 13,14] and two meta-analyses[15,16]. These meta-analyses reflect a relatively limited amount of RCT data[15] or dichotomise post-treatment improvement in a way that prevents specific estimates of effect size[16]. Given the substantial burden of disease associated

medRxiv preprint doi: https://doi.org/10.1101/2021.08.28.21262764; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

with aphasia, the present meta-analysis attempts to provide a deeper understanding about the potential and limitations of MIT. To achieve this goal, the current analyses synthesise available studies on MIT to address five research questions:

1. **Psychometric quality of outcomes.** Does the use of validated *versus* unvalidated outcomes systematically alter the resulting effect size of MIT?
2. **Experimental design.** Do RCT and non-RCT results in the context of MIT differ systematically in terms of effect size?
3. **Aphasia stage.** Do influences related to spontaneous recovery, quantified as number of months post-stroke-onset (MPO), affect the effect size of MIT?
4. **Variants of MIT protocol.** Do variations of the original MIT protocol alter the resulting effect size?
5. **Generalisability.** Apart from trained items, does MIT enhance performance on untrained items and, if so, does the resulting effect size demonstrate gains on measures of everyday communication ability?

## 2. METHODS

### 2.1. Eligibility criteria

We defined the following basic **inclusion criteria** for studies to be considered for the present meta-analysis:

- empirical study that administered MIT to adult individuals (age 18 or over) with aphasia, with or without a control group;
- language-related outcomes in pre-post assessment;
- publication in peer-reviewed journal.

medRxiv preprint doi: https://doi.org/10.1101/2021.08.28.21262764; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

We chose to include case reports with individual patient data (IPD) to increase the pool of evidence. To determine the influence of experimental design on treatment outcome, we analysed RCT and non-RCT studies separately and comparatively.

After removal of duplicate items (see section e1 in the online Supplementary Materials), the following **exclusion criteria** were applied to remaining studies:

- substantial variation from original MIT protocol[3]. We accepted *minor* changes to the MIT protocol (and examined the effect of the categorical variable: original *versus* modified MIT), as long as the protocol met all of the following features:
    - melody-based vocal expression;
    - some form of rhythmic pacing (e.g., left-hand tapping);
    - use of verbal utterances known from everyday communicative interaction;
- unvalidated outcome measures; no published or otherwise accessible validation study for the particular test battery. Exception: if a study included both trained and untrained items for an unvalidated measure, we included it to determine the degree of generalisation by comparing performance on trained and untrained items;
- other essential data not reported and / or not retrievable, even after contacting the authors (e.g., no sample size or standard error, insufficient information to compute an effect size).

The full list of included and excluded studies can be found in eTable 1 and eTable 2 (Supplementary Materials).

### 2.2. Search strategy

This was designed to obtain high search sensitivity, using both free-text and subject headings in databases, which were not restricted by language or publication form[17]. The PRISMA statement chart in Figure 1 summarises the study counts given in section e1 of the Supplementary Materials, which also documents the full literature search procedure, including search terms and databases used.

medRxiv preprint doi: https://doi.org/10.1101/2021.08.28.21262764; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

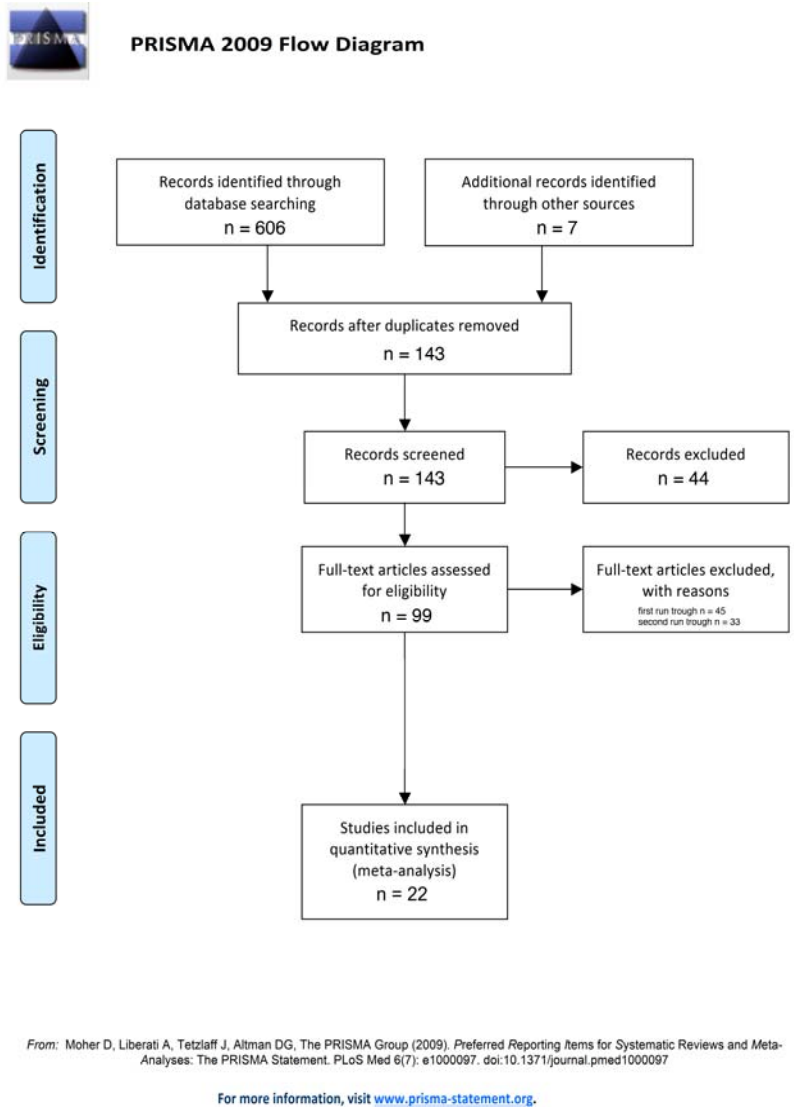



Figure 1: Flow diagram from the PRISMA statement, cf. http://prisma-statement.org/prismastatement/Checklist.aspx.

Furthermore, we followed the guidelines and standards in the Methodological Expectations of Cochrane Intervention Reviews (MECIR) handbook, and those in the PRISMA checklist (see Supplementary Materials).

### 2.3. Study coding and double-coding

All studies were coded by the first author (TP). Two of the authors (FH, TM) re-coded all studies, verifying the cross-coder consistency. Agreement among the three coders occurred in a majority of cases,

medRxiv preprint doi: https://doi.org/10.1101/2021.08.28.21262764; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

and any discrepancies found between coding sheets were solved by consensus. The ICCs (intraclass correlations) were >0.9 in the remaining cases, such as errors arisen from numerically estimating data reported in plot format only.

### 2.4. Tests and outcome measures in primary studies

eTables 3 and 4 in the Supplementary Materials respectively show: all the tests reported in the primary studies considered, and the reason for excluding some of them; a hierarchical categorisation scheme showing how, for each of the target syndromes considered (aphasia and apraxia of speech), the different measures (subtests) from the batteries of validated tests that we considered contributed towards the relevant linguistic *Abilities*, and further towards the meta-analysed dependent variables (which we deemed *Domains*).

### 2.5. MIT variants

The included MIT variants were: "SIPARI", "Music therapy", "Singing therapy", "Speech–Music Therapy for Aphasia" (SMTA), and "Modified Melodic Intonation Therapy" (MMIT). The excluded variants (cf exclusion criteria) were: choir therapy, metrical pacing technique (MPT), music therapy combined with SLT.

### 2.6. Meta-analysis methods

#### 2.6.1 Computed outcome metric

To maximise comparability of effects across studies, we used change scores from pre-test to post-test as the outcome variable, expressed in *z*-scores. For group-level studies (the RCTs in the current analyses), we standardised *z*-scores using pooled pre-test standard deviation across control and treatment groups. For individual patient data studies (the case reports in the current analyses), we computed *z*-scores in one of three ways. For studies that reported results as *z*-scores (e.g., based on test norms), we used the *z*-scores directly. For studies that reported results as percentile scores (e.g., based on test norms), we converted these to *z*-scores using the quantiles of the standard Normal distribution. For other studies, we estimated

medRxiv preprint doi: https://doi.org/10.1101/2021.08.28.21262764; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

*z*-scores using the following procedure. We first converted[a] normalised raw scores to reflect the proportion of the maximum possible score, POMP[18]. Next, we estimated a three-level random-intercept model for the pre-test POMP scores, with individual test scores nested within patients nested within studies (see Figure 2). From these models, we used the population intercept as the estimated POMP score *mean*, and the patient-level random effects standard deviation as the estimated POMP score *SD* (τ). We then used this *mean* and *SD* to standardise the pre-test and post-test POMP scores.

Figure 2: Three-level model employed, with SDs (τ) at study-, patient- and measure-level.

For models specifically fitted to the RCT and the case report data respectively, see section e4 in the Supplementary Materials.

### 2.6.2 Moderator analyses

For the RCT meta-analyses, we fit a meta-regression model with the moderators (1) Domain (cf. section 2.4); (2) whether the study used validated tests as its outcome measures, or unvalidated ones (for unvalidated measures, we treated trained and untrained items as separate groups to avoid confounding measure validation and training effects); and (3) the Domain × Validated interaction. Next, we fit another model adding the additional moderators of (1) mean MPO across treatment and control groups; and (2) the difference in mean MPO between treatment and control groups.

For the case report meta-analyses, we initially fitted the same meta-regression model with three moderators as for RCTs. We then fit two additional models adding one moderator at a time to this

[a] For a small number of studies, it was not possible to determine the maximum or minimum possible scores. For these studies, we computed POMP scores using the maximum and minimum *observed* scores in the sample. Results did not change meaningfully if we excluded these studies from results.

medRxiv preprint doi: https://doi.org/10.1101/2021.08.28.21262764; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

baseline model. First, we fit a model adding individual-level MPO. Second, we fit a model adding whether a study used the original MIT protocol or a modified protocol.

## 3. RESULTS

Study-level standardised mean difference scores and meta-analytic mean differences by Domain are shown in Figure 3. Full meta-regression results tables are reported in the Supplementary Materials, eTables 5–10.

medRxiv preprint doi: https://doi.org/10.1101/2021.08.28.21262764; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Figure 3: Results of meta-analyses. Points are study-level standardised mean pretest-posttest difference scores, either adjusted for a control group ($g_{ppc}$) or not ($g_{pp}$). Points of different colours are drawn from different studies. Large points are mean $g_{pp(c)}$ for validated measures with 66% (thick bar) and 95% (thin bar) confidence intervals and *t*-distribution confidence densities. For case reports, one aphasia severity study with $g_{ppc} = -4.88$ is not displayed.

medRxiv preprint doi: https://doi.org/10.1101/2021.08.28.21262764; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

### 3.1. RCT data

Overall, RCT data showed a small to moderate pretest-posttest effect of MIT on aphasia outcomes, after accounting for the control group ($g$ = .31 [95% CI −.01, .63]). These results were primarily based on Non-Communicative Language Expression (repetition) tasks. Other abilities were less commonly assessed. In moderator analyses, effects appeared to be much weaker for Communication and Language Comprehension tasks than for Non-Communicative Language Expression, but confidence intervals for these differences were wide (see Figure 3). Effects were estimated to be somewhat heterogeneous across studies (random effects standard deviation, $\tau$ = .33 [95% CI .15, 1.01]).

Two studies included several unvalidated measures of Non-Communicative Language Expression. For these measures, treatment effects for untrained items were somewhat smaller than those for validated measures, though the confidence interval for this difference was fairly wide ($\Delta g$ = −.15 [95% CI −.46, .15]). As expected, estimated treatment effects were much larger when patients were tested using trained items ($\Delta g$ = .99 [95% CI .60, 1.39]; trained vs. untrained items contrast: 1.15 [95% CI .74, 1.56]). Smaller effect sizes for unvalidated measures may be attributable to poorer reliability compared to validated measures; measurement error tends to attenuate effect sizes[19–21].

When aphasia stage (MPO) was added to the RCT model, neither mean MPO across groups ($\Delta g$ per month = −.008 [95% CI −.024, .008]) nor difference in mean MPO between MIT and control groups ($\Delta g$ per month = −.004 [95% CI −.020, .011]) showed meaningful relationships with MIT treatment effects. Importantly, effect sizes for RCT analyses were drawn from only three studies, so these group-level MPO analyses have limited power to estimate the impact of MPO on MIT treatment effects.

### 3.2. Case report data

Compared to RCT studies, case reports with no control group estimated much larger effects of MIT ($g$ = 1.72 [95% CI 1.00, 2.42]). As with RCT studies, these results were primarily based on Non-Communicative Language Expression (repetition) tasks. Overall aphasia severity and language comprehension appeared to show somewhat smaller effects, but confidence intervals on these differences

medRxiv preprint doi: https://doi.org/10.1101/2021.08.28.21262764; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

were very wide. Effects were estimated to be highly heterogeneous across studies ($\tau$ [between-studies] = 1.41 [95% CI .89, 2.05]), to the degree that MIT was even estimated to be harmful in a small proportion of settings: for instance, the 95% normal-theory prediction interval for Non-Communicative Language Expression ranged −0.88 to +4.90[22].

Four studies included several unvalidated measures of Non-Communicative Language Expression. As with RCT studies, treatment effects for untrained items on unvalidated measures appeared to be smaller than those for validated measures (with a wide confidence interval; $\Delta g$ = −.47 [95% CI −2.40, 1.46]). Also similar to RCTs, apparent treatment effects were much larger for trained items ($\Delta g$ = 2.37 [95% CI .44, 4.31]; trained vs. untrained items contrast: 2.84 [95% CI 1.21, 4.48]).

When aphasia stage (MPO) was added to the case reports model, MPO showed a moderate negative relationship with treatment effects ($\Delta g$ per month = −.02 [95% CI −.03, −.01]; estimated effect for 12 months, −.18 [95% CI −.30, −.07]; estimated effect for 24 months, −.37 [95% CI −.61, −.14]).

Compared to studies that used the original MIT protocol, studies that used a modified variant of the protocol appeared to show somewhat larger treatment effects, though the confidence interval on this difference was very wide ($\Delta g$ = .56 [95% CI −.92, 2.03]).

## 4. Discussion

The present meta-analysis aimed to investigate the efficacy of MIT while accounting for crucial methodological aspects of primary studies, such as validated outcomes, use of control comparison, and randomised group allocation. Our results reveal that poor methodology may introduce substantial bias into estimated treatment effects. Concerning RCT studies of non-communicative language expression, using unvalidated outcomes for untrained items may attenuate MIT's effect size by about 43% when compared to validated outcomes ($g_{unvalidated}$ = .20 vs. $g_{validated}$ = .35). Holding language domain and outcome validity constant, MIT's effect size proved to be 5.7 times larger for non-RCT data compared to

medRxiv preprint doi: https://doi.org/10.1101/2021.08.28.21262764; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

RCT data ($g_{case\ report}$ = 2.01 vs. $g_{RCT}$ = .35 for validated Non-Communicative Language Expression measures).

### 4.1. Research implications

The current results indicate that appropriate study design can help reduce confound to obtain more realistic estimates. In particular, these results re-affirm the importance of setting up and adjusting for adequate control interventions. Otherwise, most of the changes observed in case reports—visible in inflated estimates of efficacy—are inseparable from phenomena of spontaneous recovery, and ultimately, regression to the mean, none of which are due to the treatment. Effect sizes were found to decrease with number of months post-stroke for IPD studies, indicating that progress in language performance reported in the late subacute or consolidation stage of aphasia may arise from influences of spontaneous recovery. Taken together, these results suggest that validated outcomes, randomised-controlled designs and inclusion of individuals with chronic aphasia are essential prerequisites to determine the efficacy of MIT in a reliable way.

### 4.2. Clinical implications

According to the present meta-analysis, MIT leads to gains mainly in repetition tasks that reflect the ability to reproduce prior utterances in exactly the same form. Although this ability may facilitate the acquisition of novel words, it is not entirely clear to what extent it ultimately affects verbal behaviour in everyday communicative situations[23]. Our RCT results indicate negligible progress on validated outcomes of everyday communication ability with MIT. The number of non-repetition outcomes was comparatively small, regardless of experimental design, implying that benefits from MIT cannot be ruled out completely; nonetheless, current evidence does not support them. In contrast, large-scale RCT data demonstrate that combining selected non-MIT methods *can* lead to moderate gains on validated outcomes of communication ability[2]. This finding suggests that individuals with aphasia should not rely exclusively on MIT if the primary goal is to improve everyday communication. Still, our meta-analysis should not undermine the importance of MIT-mediated progress on trained items. In individuals with severe forms of

medRxiv preprint doi: https://doi.org/10.1101/2021.08.28.21262764; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

aphasia, this 'palliative' use of MIT may entail a substantial increase in quality of life[14]. Critically, individuals with aphasia may perceive notable progress in language performance irrespective of statistically significant gains on validated outcomes. Known as 'minimal clinically-important difference'[24], this diagnostic approach may be especially valuable for individuals where MIT can help establish a repertoire of trained phrases to convey basic needs in daily life[25].

### 4.3. Limitations and future directions

As with any meta-analysis, the conclusiveness of the results strongly depends on the quality of the source material. As always, methodological shortcomings of primary studies emphasise the need for caution in interpreting the results. The present meta-analysis considered various methodological aspects that tended to be neglected in previous work. In particular, our meta-analysis carefully determined the psychometric quality of each outcome, relative to recently defined standards in aphasia research[26]. In addition, our evaluation accounted for quality of the research design, in terms of using control interventions and group randomisation to address unspecific influences, including bias due to placebo effects. Our results demonstrate that the overall efficacy of MIT in repetition tasks appears to persist, albeit to a smaller degree than previously reported.

Interestingly, variation of the original MIT protocol did not systematically alter the effect size, thus challenging the idea that modification of the treatment necessarily diminishes its outcome. This finding casts doubt on the notion that the original composition and hierarchical structure of MIT are indispensable for improving language performance. However, few of the included studies employed an MIT variant, and their individual effects are heterogeneous. Therefore, our results can express no certainty about the impact of deviations from the original MIT protocol, and instead highlight the need for high-quality research exploring the influence of specific deviations.

Using unvalidated outcomes, cross-sectional and longitudinal multiple-case studies have examined the role of different MIT elements: melody and rhythm[e.g., 27], vocal expression in unison or alone[28], left-hand tapping[e.g., 29], and formulaicity of verbal utterances[e.g., 30]. Possible methodological reasons for seemingly

medRxiv preprint doi: https://doi.org/10.1101/2021.08.28.21262764; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

contradictory data, as well as conjectured mechanisms of MIT, have been discussed[e.g., 31]. Obviously, the present results do not offer insight into any of these mechanisms. If indeed adherence to the original MIT protocol does not manifest in significantly elevated language performance, our results encourage future research to optimise the composition and structure of the treatment, to increase its efficacy in the rehabilitation of neurological communication disorders. For example, individuals with apraxia of speech may benefit from several elements of MIT, such as rhythmic pacing[32] and language formulaicity[33].

### 4.4. Conclusion

We here present the first meta-analysis on MIT that attempts to rule out various methodological caveats in interpreting the outcome of previous studies, such as lack of validated outcomes, control group or randomisation. Accounting for each of these issues in a rigorous way, the results of our meta-analyses confirm the promising role of MIT in improving language performance on trained items and in repetition tasks, while highlighting possible limitations in promoting everyday communication ability. We hope that the current work will be helpful for clinicians, patients and families to make informed decisions about their treatment options to support recovery from post-stroke aphasia.

## 5. Funding and acknowledgements

This work was supported by a research-cluster grant from the Medical University of Vienna and University of Vienna (SO10300020).

We gratefully acknowledge:

- Ajay Halai and Yina Quique Buitrago for very useful comments on an early version of the manuscript.
- Joost Hurkmans for providing unpublished work.
- Sarah Wallace, Luisa Krein and Emily Braun for providing test-related information.

## 6. Author contributions

medRxiv preprint doi: https://doi.org/10.1101/2021.08.28.21262764; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

CRediT author statement:

- Conceptualization: TP, BS, WTF
- Methodology: TP, BMW
- Software: TP, BMW
- Validation: TP
- Formal analysis: TP, BMW
- Investigation: TP
- Resources: TP, RB, WTF
- Literature search and curation: TP, HH, MZ
- Data curation: TP, BMW, FH, TM
- Writing - original draft: TP, BS
- Writing - review & editing: TP, BS, BMW, FH, TM, RB, WTF
- Visualization: TP, BMW
- Supervision: TP, BS
- Project administration: TP
- Funding acquisition: RB, WTF

## 7. Competing interests

The authors declare no competing interests.

medRxiv preprint doi: https://doi.org/10.1101/2021.08.28.21262764; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

## 8. References


1. Engelter ST, Gostynski M, Papa S, et al. Epidemiology of aphasia attributable to first ischemic stroke: incidence, severity, fluency, etiology, and thrombolysis. *Stroke*. 2006;37(6):1379-1384. doi:10.1161/01.STR.0000221815.64093.8c

2. Breitenstein C, Grewe T, Flöel A, et al. Intensive speech and language therapy in patients with chronic aphasia after stroke: a randomised, open-label, blinded-endpoint, controlled trial in a health-care setting. *The Lancet*. 2017;389(10078):1528-1538. doi:10.1016/S0140-6736(17)30067-3

3. Albert ML, Sparks RW, Helm NA. Melodic Intonation Therapy for Aphasia. *Arch Neurol*. 1973;29(2):130-131. doi:10.1001/archneur.1973.00490260074018

4. Gerstman HL. A case of aphasia. *J Speech Hear Disord*. 1964;29:89-91. doi:10.1044/jshd.2901.89

5. Mills CK. Treatment of aphasia by training. *J Am Med Assoc*. 1904;43(26):1940-1949.

6. Yamadori A, Osumi Y, Masuhara S, Okubo M. Preservation of singing in Broca's aphasia. *J Neurol Neurosurg Psychiatry*. 1977;40(3):221-224. doi:10.1136/jnnp.40.3.221

7. Helm-Estabrooks N, Morgan AR, Nicholas M. *Melodic Intonation Therapy*. Pro-Ed; 1989.

8. van der Meulen I, van de Sandt-Koenderman WMiekeE, Heijenbrok-Kal MH, Visch-Brink EG, Ribbers GM. The Efficacy and Timing of Melodic Intonation Therapy in Subacute Aphasia. *Neurorehabil Neural Repair*. 2014;28(6):536-544. doi:10.1177/1545968313517753

9. Van Der Meulen I, Van De Sandt-Koenderman MWME, Heijenbrok MH, Visch-Brink E, Ribbers GM. Melodic Intonation Therapy in Chronic Aphasia: Evidence from a Pilot Randomized Controlled Trial. *Front Hum Neurosci*. 2016;10. doi:10.3389/fnhum.2016.00533

10. Doppelbauer L, Mohr B, Dreyer FR, Stahl B, Büscher V, Pulvermüller F. Long-Term Stability of Short-Term Intensive Language-Action Therapy in Chronic Aphasia: A 1-2 year Follow-Up Study. *Neurorehabil Neural Repair*. Published online July 7, 2021:15459683211029236. doi:10.1177/15459683211029235

11. RELEASE Collaborators. Predictors of Poststroke Aphasia Recovery: A Systematic Review-Informed Individual Participant Data Meta-Analysis. *Stroke*. 2021;52(5):1778-1787. doi:10.1161/STROKEAHA.120.031162

12. Blomert L, Kean ML, Koster C, Schokker J. Amsterdam—Nijmegen everyday language test: construction, reliability and validity. *Aphasiology*. 1994;8(4):381-407. doi:10.1080/02687039408248666

13. van der Meulen I, van de Sandt-Koenderman ME, Ribbers GM. Melodic Intonation Therapy: Present Controversies and Future Opportunities. *Arch Phys Med Rehabil*. 2012;93(1, Supplement):S46-S52. doi:10.1016/j.apmr.2011.05.029

14. Zumbansen A, Peretz I, Hébert S. Melodic Intonation Therapy: Back to Basics for Future Research. *Front Neurol*. 2014;5. doi:10.3389/fneur.2014.00007

medRxiv preprint doi: https://doi.org/10.1101/2021.08.28.21262764; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

15. Brady MC, Kelly H, Godwin J, Enderby P. Speech and language therapy for aphasia following stroke. *Cochrane Database Syst Rev*. 2016;(5). doi:10.1002/14651858.CD000425.pub3

16. Zumbansen A, Tremblay P. Music-based interventions for aphasia could act through a motor-speech mechanism: a systematic review and case–control analysis of published individual participant data. *Aphasiology*. 2018;33(4):466-497. doi:10.1080/02687038.2018.1506089

17. Higgins J, Thomas J, Chandler J, et al., eds. *Cochrane Handbook for Systematic Reviews of Interventions*. Cochrane; 2021. www.training.cochrane.org/handbook

18. Cohen P, Cohen J, Aiken LS, West SG. The problem of units and the circumstance for POMP. *Multivar Behav Res*. 1999;34(3):315-346. doi:10.1207/S15327906MBR3403_2

19. Ivanova MV, Hallowell B. A tutorial on aphasia test development in any language: Key substantive and psychometric considerations. *Aphasiology*. 2013;27(8):891-920. doi:10.1080/02687038.2013.805728

20. Wiernik BM, Dahlke JA. Obtaining unbiased results in meta-analysis: the importance of correcting for statistical artifacts. *Adv Methods Pract Psychol Sci*. 2020;3(1):94-123. doi:10.1177/2515245919885611

21. van Smeden M, Lash TL, Groenwold RHH. Reflection on modern methods: five myths about measurement error in epidemiological research. *Int J Epidemiol*. 2020;49(1):338-347. doi:10.1093/ije/dyz251

22. IntHout J, Ioannidis JPA, Rovers MM, Goeman JJ. Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open*. 2016;6(7):e010247. doi:10.1136/bmjopen-2015-010247

23. Stahl B, Mohr B, Dreyer FR, Lucchese G, Pulvermüller F. Communicative-Pragmatic Assessment Is Sensitive and Time-Effective in Measuring the Outcome of Aphasia Therapy. *Front Hum Neurosci*. 2017;11. doi:10.3389/fnhum.2017.00223

24. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol*. 2008;61(2):102-109. doi:10.1016/j.jclinepi.2007.03.012

25. Van Lancker Sidtis D. *Foundations of Familiar Language: Formulaic Expressions, Lexical Bundles, and Collocations at Work and Play*. 1st edition. Wiley-Blackwell; 2021.

26. Wallace SJ, Worrall L, Rose T, et al. A core outcome set for aphasia treatment research: The ROMA consensus statement. *Int J Stroke*. 2019;14(2):180-185. doi:10.1177/1747493018806200

27. Kershenbaum A, Nicholas ML, Hunsaker E, Zipse L. Speak along without the song: what promotes fluency in people with aphasia? *Aphasiology*. 2019;33(4):405-428. doi:10.1080/02687038.2017.1413487

28. Racette A, Bard C, Peretz I. Making non-fluent aphasics speak: sing along! *Brain*. 2006;129(10):2571-2584. doi:10.1093/brain/awl250

medRxiv preprint doi: https://doi.org/10.1101/2021.08.28.21262764; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

29. Zipse L, Worek A, Guarino AJ, Shattuck-Hufnagel S. Tapped out: do people with aphasia have rhythm processing deficits? *J Speech Lang Hear Res JSLHR*. 2014;57(6):2234-2245. doi:10.1044/2014_JSLHR-L-13-0309

30. Stahl B, Kotz SA, Henseler I, Turner R, Geyer S. Rhythm in disguise: why singing may not hold the key to recovery from aphasia. *Brain*. 2011;134(10):3083-3093. doi:10.1093/brain/awr240

31. Stahl B, Kotz SA. Facing the music: three issues in current research on singing and aphasia. *Front Psychol*. 2014;5:1033. doi:10.3389/fpsyg.2014.01033

32. Aichert I, Späth M, Ziegler W. The role of metrical information in apraxia of speech. Perceptual and acoustic analyses of word stress. *Neuropsychologia*. 2016;82:171-178. doi:https://doi.org/10.1016/j.neuropsychologia.2016.01.009

33. Stahl B, Gawron B, Regenbrecht F, Flöel A, Kotz SA. Formulaic Language Resources May Help Overcome Difficulties in Speech-Motor Planning after Stroke. *PloS One*. 2020;15(6):e0233608. doi:10.1371/journal.pone.0233608

medRxiv preprint doi: https://doi.org/10.1101/2021.08.28.21262764; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

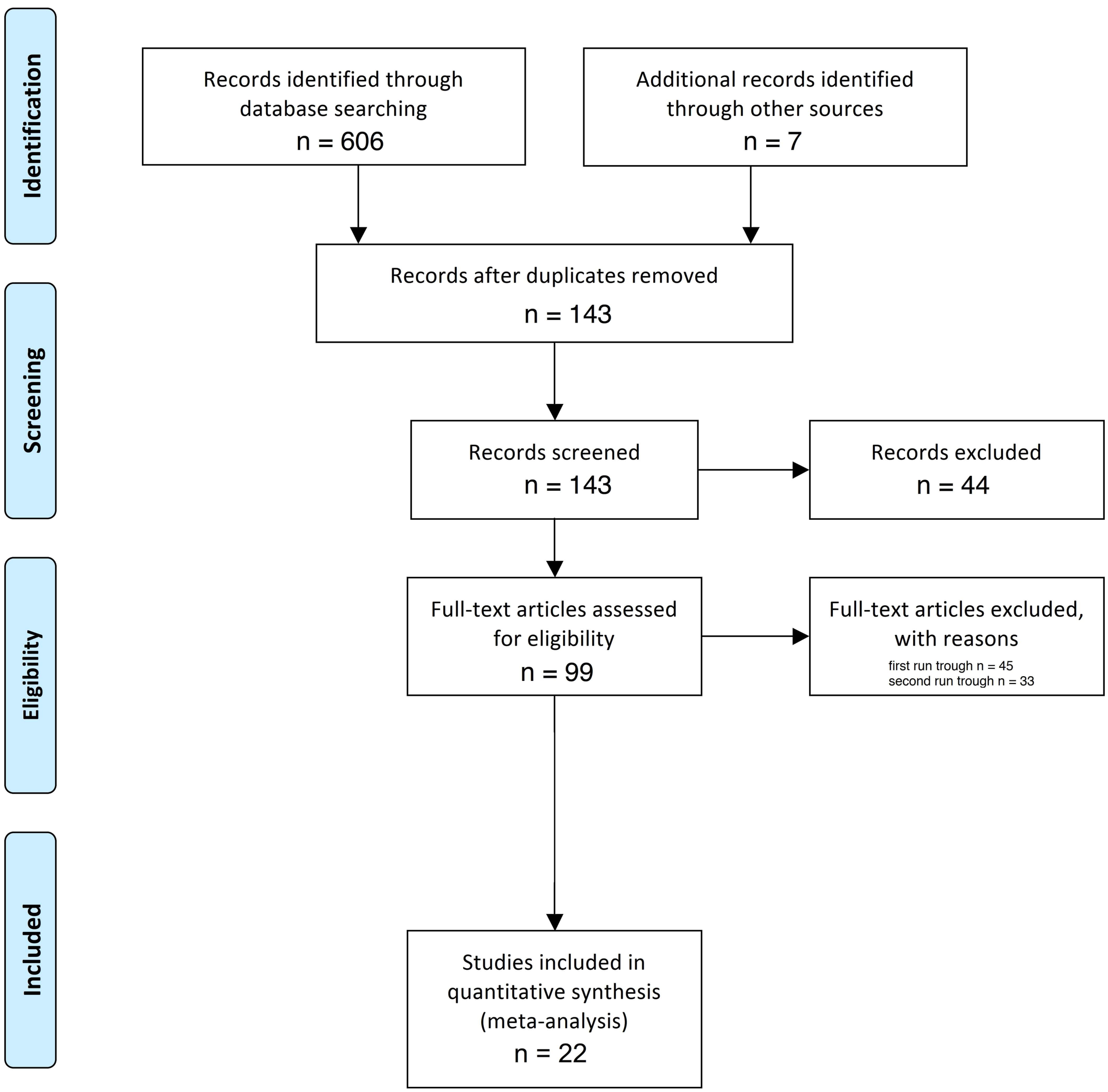
## 9. TABLE OF FIGURES

Figure 1: Flow diagram from the PRISMA statement, cf. http://prisma-statement.org/prismastatement/Checklist.aspx. ..... 7

Figure 2: Three-level model employed, with SDs (τ) at study-, patient- and measure-level. ..... 9

Figure 3: Results of meta-analyses. Points are study-level standardised mean pretest-posttest difference scores, either adjusted for a control group ($g_{ppc}$) or not ($g_{pp}$). Points of different colours are drawn from different studies. Large points are mean $g_{pp(c)}$ for validated measures with 66% (thick bar) and 95% (thin bar) confidence intervals and *t*-distribution confidence densities. For case reports, one aphasia severity study with $g_{ppc} = -4.88$ is not displayed. ..... 11

# PRISMA 2009 Flow Diagram



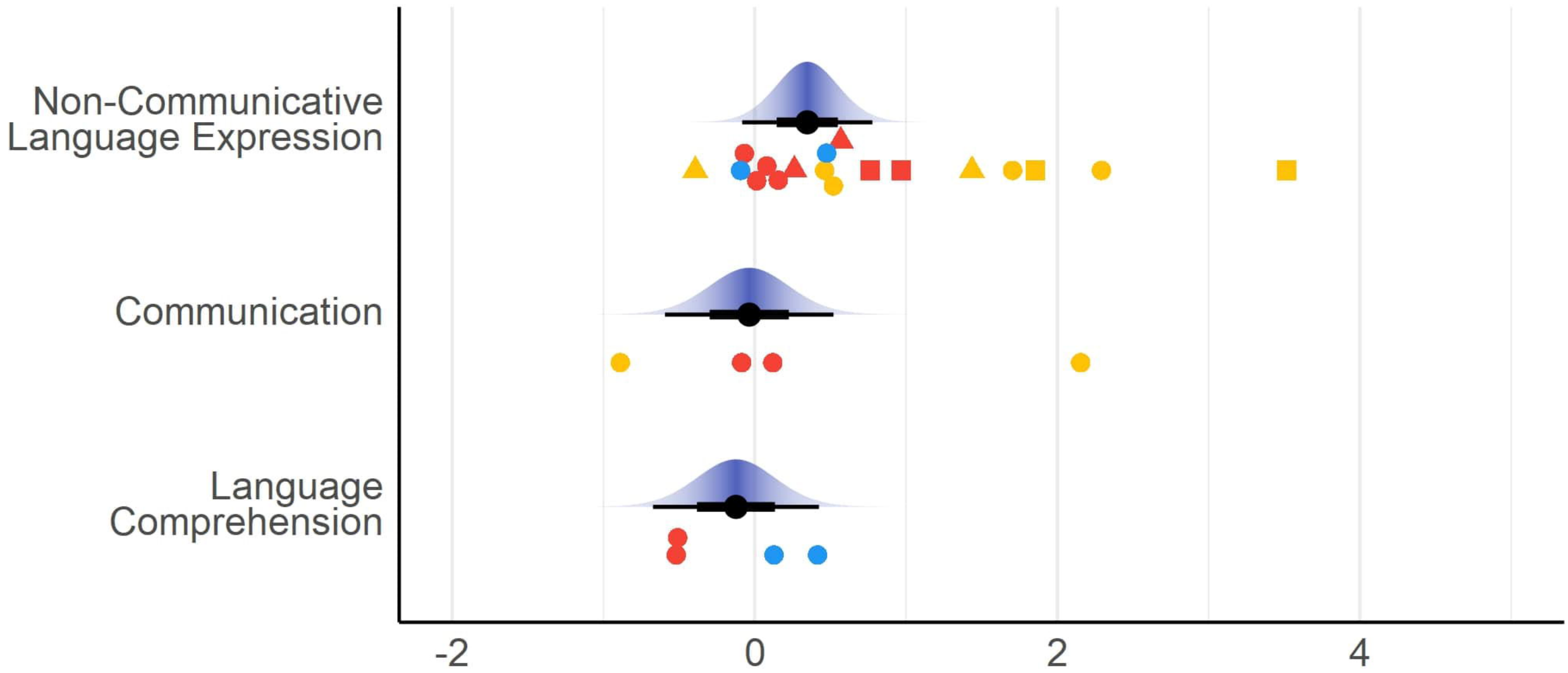

modified preprint doi: https://doi.org/10.1101/2021.08.25.21262764; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

*From:* Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). *P*referred *R*eporting *I*tems for *S*ystematic Reviews and *M*eta-*A*nalyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

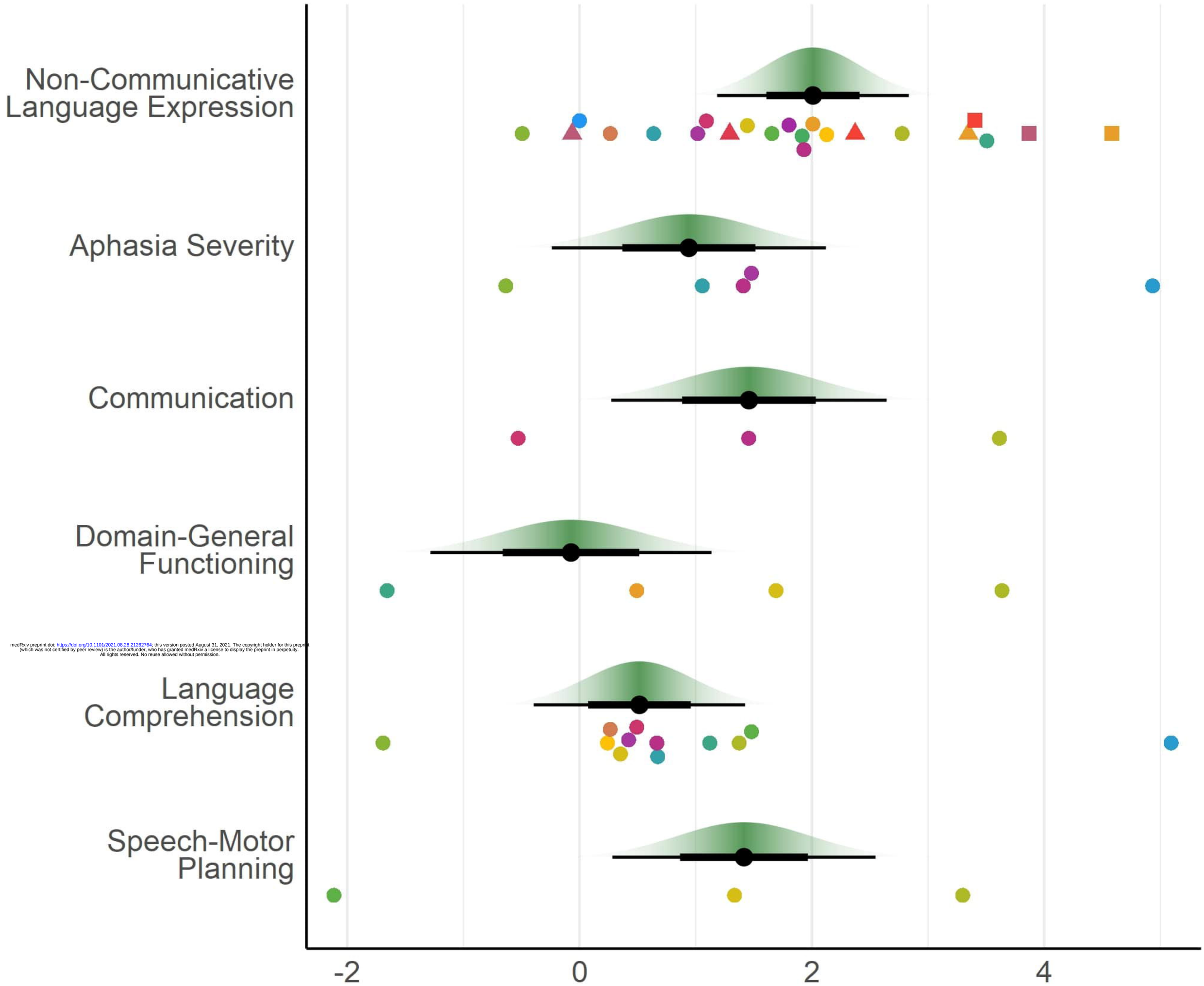
**For more information, visit www.prisma-statement.org.**

τstudies
τpatients
τmeasures

## Randomised Control Trials: $g_{ppc}$ by Domain




## Case Reports: $g_{pp}$ by Domain




medRxiv preprint doi: https://doi.org/10.1101/2021.08.28.21262764; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

● Validated ▲ Unvalidated (untrained) ■ Unvalidated (trained)