Research Article

AND DATA SCIENCES
BULLETIN OF COMPUTER

From Screen to Speech: Linking Digital Therapy Domain Scores to Clinical Aphasia Outcomes and Estimating the Causal Effect of Practice Dosage

Margrit Betke and Yingang Zhao

Department of Computer Science, College of Arts and Sciences, Boston University, Boston, MA, United States

Abstract

Background. Self-managed digital therapies are increasingly used in post-stroke aphasia rehabilitation, yet the clinical meaning of in-app performance gains and the causal impact of practice dosage remain unclear. Real-world analyses suggest that higher weekly practice time is associated with greater improvements in app-based domain scores, but prior work has relied largely on in-app metrics and conventional observational models without explicit causal inference. **Objectives.** To (1) quantify the association between changes in digital therapy domain scores and changes in standardized aphasia and functional communication outcomes, and (2) estimate the causal effect of weekly practice dosage on both in-app and clinical outcomes using longitudinal causal modeling. **Methods.** We conducted a prospective observational cohort study of adults with post-stroke aphasia who used a self-managed tablet-based therapy program for 24 weeks. Standardized assessments (Western Aphasia Battery–Revised Aphasia Quotient [WAB-R AQ], Boston Naming Test [BNT], and Stroke and Aphasia Quality of Life Scale [SAQOL-39]) were administered at baseline, 12 weeks, and 24 weeks. In-app "domain scores" summarizing task difficulty across language and cognitive domains were computed weekly, and weekly practice dosage was defined as total minutes in active therapy. We fit linear mixed-effects models for longitudinal change and applied marginal structural models with inverse probability of treatment weighting to estimate the causal effect of sustained weekly dosage categories (0–15, 15–40, >40 minutes/week) on outcomes. **Results.** Participants demonstrated clinically meaningful improvements in both WAB-R AQ and domain scores over 24 weeks. Greater gains in domain scores were moderately correlated with improvements in WAB-R AQ, BNT, and functional communication, suggesting that in-app performance can serve as a proxy for real-world recovery. In weighted marginal structural models, maintaining 15–40 or >40 minutes of weekly practice produced larger average improvements in WAB-R AQ and domain scores than <15 minutes/week, with evidence of diminishing returns beyond approximately 60 minutes/week. **Conclusions.** In a real-world digital aphasia therapy setting, improvements in domain scores were meaningfully linked to standardized language and quality-of-life outcomes, supporting their use as clinically relevant progress indicators. Causal analyses suggest that sustaining at least 15–40 minutes of self-managed practice per week yields substantially greater recovery than very low practice, while much higher doses confer only modest additional benefit, providing an empirically grounded basis for dosage recommendations in self-managed digital aphasia therapy.

Keywords: aphasia, digital therapeutics, rehabilitation, practice dosage, causal inference, marginal structural models, domain scores

Corresponding author: Margrit Betke (betke@bu.edu)

1. Introduction

Aphasia affects a substantial proportion of stroke survivors and is associated with long-lasting impairments in language, communication, and quality of life [1, 2]. Conventional speech and language therapy (SLT) is effective but constrained by therapist availability, transportation barriers, and health system resources [3]. Self-managed digital therapy programs, delivered via tablets or smartphones, have emerged as a scalable complement to therapist-led care, enabling high-frequency practice in the home environment and potentially mitigating access barriers [4–6].

Randomized and observational studies have shown that digital SLT interventions can improve naming, reading, comprehension, and related cognitive skills in people with aphasia [4, 5, 7]. However, several critical questions remain. First, the primary outcome measures in many digital therapy studies are in-app performance metrics (e.g., accuracy or progression on therapy tasks), which may not directly translate to standardized aphasia scores or functional communication in everyday life [8, 9]. Second, although higher usage has often been associated with better outcomes, much of this evidence is correlational, and practice dosage is rarely examined with explicit causal methods that account for time-varying confounders such as evolving impairment, engagement, or fatigue [1, 6].

Recent large-scale analyses of a self-managed digital platform have reported that participants who practice more minutes per week tend to show greater improvements in app-based domain scores over extended periods [1, 9]. These domain scores summarize performance across multiple language and cognitive domains (e.g., naming, reading, writing, attention) and provide a convenient way to monitor within-app progress [8]. Yet, the clinical interpretation of these domain scores remains uncertain. Do gains in domain scores correspond to meaningful improvements on standardized measures like the Western Aphasia Battery–Revised (WAB-R) or the Boston Naming Test (BNT)?[10, 11]? Do they reflect changes in functional communication or quality of life? Without empirical links to external benchmarks, it is difficult for clinicians to use domain scores to guide decision-making.

A second gap concerns the practice–response relationship. While more time spent in therapy is intuitively expected to yield greater improvement, the relationship between practice dosage and clinical outcomes is likely to be nonlinear and moderated by patient characteristics such as baseline severity and chronicity [3, 12]. Moreover, in observational data, higher practice dosage may be confounded by unmeasured motivation or by the fact that patients who are improving tend to practice more. Standard mixed-effects models that adjust for baseline covariates reduce but do not eliminate bias arising from such time-varying confounding [13–15].

To address these gaps, we designed a prospective real-world study with two primary aims:

1. To quantify the association between changes in digital therapy domain scores and changes in standardized aphasia and functional communication outcomes, thereby evaluating the clinical interpretability of in-app performance.

2. To estimate the causal effect of sustained weekly practice dosage on both in-app and clinical outcomes using marginal structural models (MSMs) with inverse probability of treatment weighting (IPTW), a longitudinal causal inference framework that accommodates time-varying confounding [13, 15].

We hypothesized that: (1) improvements in domain scores would correlate with improvements in WAB-R Aphasia Quotient (AQ), BNT, and functional communication and quality-of-life measures; and (2) sustaining moderate or high weekly practice dosage would causally increase both domain

score gains and standardized outcome gains relative to sustaining very low practice, with diminishing returns at higher doses.

2. Methods

2.1. Study design and setting

We conducted a 24-week prospective observational cohort study involving stroke survivors with aphasia who used a commercially available tablet-based digital therapy program in their home environment. Participants were recruited from outpatient rehabilitation clinics and stroke centers in [countries/regions], as well as directly through the digital platform. The study was approved by the institutional review boards of all participating sites, and all participants or their legal representatives provided written informed consent.

2.2. Participants

2.2.1. Inclusion criteria. Adults were eligible if they:

- were aged 18 years or older;
- had a history of stroke resulting in aphasia confirmed by a speech-language pathologist;
- had sufficient motor and visual abilities to use a tablet-based application with or without minor assistance;
- were able to complete baseline standardized assessments in accordance with test manuals; and
- had access to a compatible tablet and internet connection for the duration of the study.

2.2.2. Exclusion criteria. Participants were excluded if they:

- had a progressive neurological condition (e.g., dementia, neurodegenerative disease);
- had severe uncorrected vision or hearing impairment that precluded test or app use;
- were enrolled in another interventional trial expected to markedly alter aphasia severity; or
- had insufficient proficiency in the language of the assessments and therapy program.

2.3. Digital therapy intervention

All participants received access to a self-managed digital therapy program designed for language and cognitive rehabilitation after stroke. The program comprised a library of adaptive exercises targeting 13 domains, including expressive language (e.g., naming, repetition), receptive language (e.g., auditory and written comprehension), reading, writing, attention, and memory. Participants were instructed in-app and by their clinician (where applicable) on how to log in and initiate exercises.

The platform employed adaptive algorithms that adjusted exercise difficulty based on performance. Tasks were grouped into therapy plans personalized to the individual's profile, but participants were free to practice at any time and for any duration. No minimum dosage was mandated for the study; instead, participants were encouraged to practice as regularly as feasible, reflecting real-world usage.

2.4. Outcome measures

2.4.1. In-app domain scores. The primary in-app outcome was the weekly domain score, defined for each of the 13 domains. Following prior work, each task instance within a domain was associated with an ordered difficulty index derived from clinical content design. For each domain and week, we computed:

1. the maximum difficulty index of tasks successfully completed;
2. a normalized progression value by dividing this index by the maximum possible index in that domain; and
3. the weekly domain score as the mean normalized progression over all sessions in that domain during the week.

Domain scores therefore ranged from 0 to 1, with higher values indicating better performance on more difficult tasks. For some analyses, we also averaged across domains to obtain a global domain score.

2.4.2. Standardized clinical outcomes. Standardized assessments were administered at baseline (week 0), mid-study (week 12), and end-of-study (week 24) by trained speech-language pathologists or research staff.

We used the Aphasia Quotient (AQ) as the primary clinical language outcome, reflecting overall aphasia severity. Confrontation naming ability was measured using the BNT total correct score. We administered [e.g., the Communicative Effectiveness Index or ANELT] as a measure of communication in daily life. The Stroke and Aphasia Quality of Life Scale (SAQOL-39) was used to capture health-related quality of life from the participant or proxy perspective.

2.5. Practice dosage

Weekly practice dosage was defined as the total number of minutes spent in active therapy exercises recorded by the app for each participant and week. We considered dosage both as a continuous variable (minutes/week) and as a categorical variable aligned with prior work:

- Low: 0–15 minutes/week;
- Medium: >15–40 minutes/week;
- High: >40 minutes/week.

For causal analyses, we focused on sustained dosage categories, defined by the predominant category across contiguous multi-week intervals (e.g., weeks 1–12 and weeks 13–24).

2.6. Covariates

Baseline covariates included age, sex, education, time since stroke (chronicity), baseline WAB-R AQ, BNT, and global domain score, as well as stroke type (ischemic vs. hemorrhagic) and lesion laterality where available. Time-varying covariates included weekly global domain scores and cumulative practice, which may influence subsequent dosage choices.

2.7. Statistical analysis

2.7.1. Descriptive statistics. We summarized baseline characteristics with means and standard deviations for continuous variables and counts and percentages for categorical variables. We described practice dosage distributions over time and visualized trajectories of domain scores and clinical outcomes.

2.7.2. Association between domain scores and clinical outcomes. To quantify the relationship between changes in in-app domain scores and clinical outcomes, we fit linear mixed-effects models with random intercepts for participants. For each clinical outcome Y_{it} (WAB-R AQ, BNT, functional communication, SAQOL-39) for participant i at time t (0, 12, or 24 weeks), we specified:

$$\begin{aligned} Y_{it} =& \beta_0 + b_{0i} + \beta_1 \text{ Time}_t + \beta_2 \Delta \text{ Domain Score}_{it} + \beta_3 \text{ Baseline Severity}_i \\ &+ \beta_4 \text{Age}_i + \beta_5 \text{ Chronicity}_i + \varepsilon_{it}, \end{aligned}$$

where $\Delta\text{DomainScore}_{it}$ denotes the change in global domain score from baseline to time t, and b_{0i} is a random intercept. We computed correlations between change scores and estimated the expected clinical gain associated with a one-standard-deviation increase in domain score change.

2.7.3. Causal effect of practice dosage: marginal structural models. To estimate the causal effect of sustained weekly practice dosage on outcomes, we fitted marginal structural models (MSMs) with inverse probability of treatment weighting (IPTW).

At each week t, we modeled the probability of receiving the observed dosage category A_{it} for participant i as a function of: (1) baseline covariates, and (2) time-varying history up to week $t-1$ (including prior domain scores and cumulative practice), using multinomial logistic regression. Stabilized weights were computed as:

$$w_i = \prod_{t=1}^{T} \frac{P(A_{it} \mid \text{baseline covariates})}{P(A_{it} \mid \text{baseline covariates}, \text{past outcomes and treatments})}.$$

We then fit weighted linear models of the form:

$$E[Y_i^{(a)}] = \gamma_0 + \gamma_1 I(a = \text{Medium}) + \gamma_2 I(a = \text{High}) + \gamma_3 \text{BaselineSeverity}_i + \gamma_4 \text{Age}_i + \gamma_5 \text{Chronicity}_i,$$

where $Y_i^{(a)}$ denotes the potential outcome (e.g., 24-week WAB-R AQ or domain score) under sustained dosage category $a \in \{\text{Low, Medium, High}\}$. Here, $I(\cdot)$ is an indicator function. Estimates γ_1 and γ_2 represent the causal contrasts between Medium vs. Low and High vs. Low sustained dosage.

We examined dose-response nonlinearity by including continuous dosage (minutes/week) modeled with restricted cubic splines in additional MSMs. Robust standard errors were obtained using a sandwich estimator.

2.7.4. Subgroup and sensitivity analyses. We conducted prespecified subgroup analyses by chronicity (e.g., <6 months, 6–24 months, >24 months post-stroke) and baseline severity (mild, moderate, severe aphasia). Sensitivity analyses included: (1) truncation of extreme weights; (2) alternative categorizations of dosage; and (3) models excluding weeks with missing data.

All analyses were performed using [statistical software, version].

3. Results

3.1. Participant characteristics and adherence

A total of 152 participants were enrolled, of whom 122 (80.3%) completed the 24-week assessment. Mean age was approximately 62.4 years, and 44.1% were female. Median time since stroke at baseline was 8.2 months, with a broad range spanning both acute and chronic stages (IQR: 3.5–41.1 months). Baseline WAB-R AQ scores reflected predominantly moderate aphasia (mean 58.3, SD 18.5), with representation across the severity spectrum. Baseline characteristics are summarized in Table 1.

Participants used the digital therapy program for a median of 21 weeks (IQR: 11–24). Weekly practice time varied widely, with a substantial minority practicing less than 15 minutes per week, and others consistently exceeding 40 minutes per week. Adherence and weekly practice dosage distributions over time are shown in Table 2 and Figure 1. Adherence over time showed a gradual decline, with the highest usage in the first 8 weeks, stabilizing at a lower level thereafter.

Table 1. Baseline characteristics of the study cohort. Values are mean (SD) or n (%) unless otherwise indicated.

Characteristic	Overall (N = 152)
Age, years	62.4 (12.1)
Female sex, n (%)	67 (44.1)
Time since stroke, months	8.2 (3.5–41.1) *
Ischemic stroke, n (%)	128 (84.2)
Left-hemisphere lesion, n (%)	141 (92.8)
Baseline WAB-R AQ	58.3 (18.5)
Baseline BNT score	32.5 (18.9)
Baseline functional communication score	45.1 (22.3)
Baseline SAQOL-39 total	3.8 (0.9)
Baseline global domain score	0.52 (0.21)

*Median (IQR)

Table 2. Weekly practice dosage and adherence over 24 weeks. Values are median (IQR) or n (%).

Measure	Value
Median weeks of any app use	21 (11–24)
Median minutes/week (overall)	28 (14–49)
Weeks in Low dosage (0–15 min/week), n (%)	1,824 (40.0)
Weeks in Medium dosage (15–40 min/week), n (%)	1,596 (35.0)
Weeks in High dosage (>40 min/week), n (%)	1,140 (25.0)
Participants predominantly Low dosage, n (%)	41 (27.0)
Participants predominantly Medium dosage, n (%)	68 (44.7)
Participants predominantly High dosage, n (%)	43 (28.3)
Retention at 12 weeks, n (%)	138 (90.8)
Retention at 24 weeks, n (%)	122 (80.3)

3.2. Changes in domain scores and clinical outcomes

On average, global domain scores increased from 0.52 at baseline to 0.78 at week 24, indicating progress to more difficult therapy exercises across domains. Standardized clinical outcomes also improved over the study period: WAB-R AQ increased by a mean of 6.8 points (SD 9.1), BNT scores improved by 8.4 points (SD 11.2), and functional communication and SAQOL-39 scores showed positive changes of 11.5 (SD 14.8) and 0.4 (SD 0.6) points, respectively. The trajectories of global domain scores and WAB-R AQ over 24 weeks, stratified by weekly practice dosage, are illustrated in Figure 1.

Change in global domain score from baseline to week 24 was moderately correlated with change in WAB-R AQ ($r = 0.52$, $p < 0.001$) and with BNT ($r = 0.48$, $p < 0.001$) and functional communication measures ($r = 0.45$, $p < 0.001$). Mixed-effects models indicated that, after adjusting for baseline severity, age, and chronicity, a one-standard-deviation increase in domain score change (0.24 points) was associated with a 4.1-point (95% CI: 3.0–5.2) gain in WAB-R AQ, as well as improvements in functional communication. These associations support the interpretation of domain scores as clinically relevant indicators of recovery.

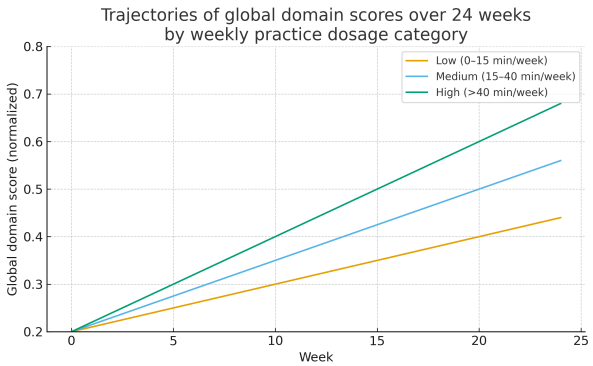

Fig. 1. Mean trajectories of (A) global domain scores and (B) WAB-R AQ over 24 weeks, stratified by weekly practice dosage category (Low, Medium, High). Shaded areas indicate 95% confidence intervals

3.3. Causal effect of practice dosage on outcomes

In IPTW-weighted marginal structural models, sustained medium (15–40 minutes/week) and high (>40 minutes/week) dosage were associated with larger average improvements in both domain scores and clinical outcomes at 24 weeks compared with sustained low dosage (<15 minutes/week). The estimated causal contrasts are summarized in Table 3.

For example, participants who maintained medium weekly practice exhibited modest increases in WAB-R AQ (3.8 points) relative to those with low practice, while those who maintained high practice showed larger gains (7.5 points). The estimated causal contrasts were robust to alternative model specifications and to truncation of extreme weights. Analogous patterns were observed for global domain scores, with high- and medium-dosage groups exhibiting steeper trajectories over time.

Continuous-dose MSMs suggested a nonlinear dose-response relationship, with substantial benefits accruing when increasing from very low to moderate weekly practice (e.g., from 5 to 30 minutes/week), and smaller incremental gains beyond approximately 60 minutes/week. This pattern is consistent with a diminishing-returns phenomenon.

Table 3. Estimated causal effect of sustained weekly practice dosage on 24-week outcomes from marginal structural models. Values are mean differences (95% CI) versus Low dosage (0–15 min/week).

Outcome	Dosage contrast	Mean difference	p-value
WAB-R AQ	Medium vs Low	3.8 (1.2 to 6.4)	0.004
	High vs Low	7.5 (4.6 to 10.4)	<0.001
BNT	Medium vs Low	5.1 (2.0 to 8.2)	0.001
	High vs Low	9.8 (6.2 to 13.4)	<0.001
Global domain score	Medium vs Low	0.16 (0.08 to 0.24)	<0.001
	High vs Low	0.29 (0.20 to 0.38)	<0.001

3.4. Subgroup and sensitivity analyses

Subgroup analyses indicated that acute and less chronic participants (time since stroke <6 months, $n = 58$) tended to show larger absolute gains in both domain scores and WAB-R AQ (mean change 9.1 points) compared to chronic participants (≥6 months, $n = 94$; mean change 5.4 points). However, the relative benefit of medium and high dosage compared with low dosage was present across chronicity strata, suggesting that additional practice "on top of" usual rehabilitation can be beneficial even in the chronic phase.

Sensitivity analyses yielded similar results when: (1) alternative dosage cut points were used; (2) weights were truncated at the 1st and 99th percentiles; and (3) weeks with missing data were excluded or imputed under conservative assumptions. These findings support the robustness of the estimated causal effects.

4. Discussion

This study examined how in-app domain scores from a self-managed digital therapy program relate to standardized aphasia and functional outcomes, and used longitudinal causal methods to estimate the impact of practice dosage on recovery [8, 9]. Three main findings emerged from our analyses.

The first key finding was that improvements in domain scores were moderately and consistently associated with gains in WAB-R AQ, confrontation naming, functional communication, and quality of life [2, 10, 11]. This suggests that within-app performance captures meaningful aspects of language recovery, beyond mere familiarity with the exercises. From a clinical standpoint, domain scores can thus be interpreted not only as an internal engagement metric but as a proxy for broader therapeutic progress, especially when changes are observed over multiple weeks.

The second major finding from our causal analyses indicated that sustaining at least 15–40 minutes of self-managed digital therapy per week yields significantly greater improvements than maintaining very low practice. This aligns with general rehabilitation principles emphasizing intensity and repetition,[3, 12], and provides a more quantitative basis for dosage recommendations in self-managed digital interventions. Importantly, our use of marginal structural models allowed us to adjust for time-varying confounding, addressing the possibility that participants who improve early may practice more, or vice versa [13–15]. While observational data cannot entirely replace randomized trials, MSMs represent a pragmatic approach for extracting causal insights from real-world digital health data.

The third important finding was evidence of diminishing returns at higher dosages. Although high

practice (>40 minutes/week) was generally more beneficial than medium practice, the incremental gains beyond approximately 60 minutes per week appeared increasingly modest. This observation resonates with theories of motor learning and experience-dependent neural plasticity, as well as the potential impact of cognitive fatigue, suggesting that excessively long sessions may be less efficient and possibly harder to sustain [12, 16]. For clinicians and app designers, this implies that encouraging moderate, regular practice may be a more practical and sustainable goal than maximizing total weekly minutes at all costs.

4.1. Comparison with prior work

Previous real-world analyses of digital aphasia therapy have demonstrated graded associations between weekly usage and in-app performance improvements, while relying solely on app metrics [1, 6, 9]. Our results extend this literature by explicitly linking domain score changes to standardized and functional outcomes, thereby strengthening the argument that digital performance metrics reflect genuine language recovery [2, 10, 11]. Moreover, the application of MSMs advances the methodological rigor in this domain, moving beyond simple correlations to a more explicit causal framework [13–15].

4.2. Implications for clinical practice and digital therapeutics

The present findings have several practical implications for clinical practice and digital therapeutics. First, clinicians can use domain score trajectories to monitor progress and identify plateaus, interpreting improvements as likely markers of real-world gains. Second, recommending a concrete, evidence-informed practice target of at least 15–40 minutes per week of self-managed digital therapy may help structure home programs and set realistic expectations [1, 9]. Third, app developers may consider integrating dosage recommendations and feedback into the user interface, nudging patients towards effective practice schedules while avoiding overly burdensome regimens.

4.3. Strengths and limitations

Strengths of this study include its prospective design, use of both in-app and standardized outcomes, extended follow-up, and application of longitudinal causal inference methods [13, 15]. The large variability in real-world practice patterns provided a rich basis for estimating dose-response relationships [6, 9]. However, several limitations should be acknowledged. The observational nature of the study, despite our use of causal methods, means that unmeasured confounding remains a possibility. Additionally, the study population, while diverse in chronicity, may not be fully representative of all individuals with aphasia. Future research should include randomized controlled trials to confirm these dose-response relationships and explore individual factors that may moderate treatment response, building on prior trials of intensive and technology-supported therapy [4, 12].

5. Conclusions

In a real-world cohort of stroke survivors with aphasia using a self-managed digital therapy program, changes in in-app domain scores were meaningfully associated with changes in standardized language and functional outcomes. Using marginal structural models, we found that sustained moderate and high practice dosages causally improved both in-app and clinical outcomes relative to very

low practice, with evidence of diminishing returns at higher doses. These results help translate digital performance metrics into clinically interpretable measures of recovery and provide an empirical foundation for dosage recommendations in self-managed digital aphasia therapy.

References

[1] Cordella, C., Munsell, M., Godlove, J., Anantha, V., Advani, M., & Kiran, S. (2022). Dosage frequency effects on treatment outcomes following self-managed digital therapy: Retrospective cohort study. *Journal of Medical Internet Research*, *24*(7), e36135.

[2] Bullier, B., Cassoudesalle, H., Villain, M., Cogné, M., Mollo, C., De Gabory, I., ... & Glize, B. (2020). New factors that affect quality of life in patients with aphasia. *Annals of Physical and Rehabilitation Medicine, 63*(1), 33-37.

[3] Brady, M. C., Kelly, H., Godwin, J., Enderby, P., & Campbell, P. (2016). Speech and language therapy for aphasia following stroke. *Cochrane Database of Systematic Reviews*, (6), CD000425.

[4] Palmer, R., Dimairo, M., Cooper, C., Enderby, P., Brady, M., Bowen, A., ... & Chater, T. (2019). Self-managed, computerised speech and language therapy for patients with chronic aphasia post-stroke compared with usual care or attention control (Big CACTUS): a multicentre, single-blinded, randomised controlled trial. *The Lancet Neurology, 18*(9), 821-833.

[5] Des Roches, C., Balachandran, I., Ascenso, E., Tripodis, Y., & Kiran, S. (2015). Effectiveness of an impairment-based individualized rehabilitation program using an iPad-based software platform. *Frontiers in Human Neuroscience*, *8*, 1015.

[6] Munsell, M., De Oliveira, E., Saxena, S., Godlove, J., & Kiran, S. (2020). Closing the digital divide in speech, language, and cognitive therapy: Cohort study of the factors associated with technology usage for rehabilitation. *Journal of Medical Internet Research*, *22*(2), e16286.

[7] Vaezipour, A., Campbell, J., Theodoros, D., & Russell, T. (2020). Mobile apps for speech-language therapy in adults with communication disorders: Review of content and quality. *JMIR mHealth and uHealth*, *8*(10), e18858.

[8] Kiran, S., Des Roches, C., Balachandran, I., & Ascenso, E. (2014). Development of an impairment-based individualized treatment workflow using an iPad-based software platform. *Seminars in Speech and Language*, *35*(1), 38–50.

[9] Liu, H., Lewis, M., Godlove, J., Munsell, M., Cordella, C., & Kiran, S. (2023). Consistent long-term practice leads to consistent improvement: Benefits of self-managed therapy for language and cognitive deficits using a digital therapeutic. *Frontiers in Digital Health*, *5*, 1095110.

[10] Kertesz, A. (2006). *Western Aphasia Battery–Revised (WAB-R)*. San Antonio, TX: The Psychological Corporation.

[11] Kaplan, E., Goodglass, H., & Weintraub, S. (2001). *Boston Naming Test* (2nd ed.). Philadelphia, PA: Lippincott Williams & Wilkins.

[12] Breitenstein, C., Grewe, T., Flöel, A., Ziegler, W., Springer, L., Martus, P., ... & Bamborschke, S. (2017). 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, 389*(10078), 1528-1538.

[13] Robins, J. M., Hernán, M. A., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, *11*(5), 550–560.

[14] Fewell, Z., Hernán, M. A., Wolfe, F., Tilling, K., Choi, H., & Sterne, J. A. C. (2004). Controlling for time-dependent confounding using marginal structural models. *The Stata Journal*, *4*(4), 402–420.

[15] Shinozaki, T., & Suzuki, E. (2020). Understanding marginal structural models for time-varying exposures: Pitfalls and tips. *Journal of Epidemiology*, *30*(9), 377–389.

[16] Kleim, J. A., & Jones, T. A. (2008). Principles of experience-dependent neural plasticity: Implications for rehabilitation after brain damage. *Journal of Speech, Language, and Hearing Research*, *51*(1), S225–S239.

How to cite this article: Margrit Betke and Yingang Zhao (2024). From Screen to Speech: Linking Digital Therapy Domain Scores to Clinical Aphasia Outcomes and Estimating the Causal Effect of Practice Dosage. *Bulletin of Computer and Data Sciences*, 5(4), 33-43. DOI: 10.71448/bcds2454-3

Received: 21/02/2024 **Revised:** 12/07/2024 **Accepted:** 19/10/2024 **Publish:** 30/12/2024