

Supplement

Description of an Intensive Residential Aphasia Treatment Program: Rationale, Clinical Processes, and Outcomes

Ronda L. Winans-Mitrik,^a William D. Hula,^{a,b} Michael W. Dickey,^{a,b}
James G. Schumacher,^a Brooke Swoyer,^a and Patrick J. Doyle^{a,b}

Purpose: The purpose of this article is to describe the rationale, clinical processes, and outcomes of an intensive comprehensive aphasia program (ICAP).

Method: Seventy-three community-dwelling adults with aphasia completed a residentially based ICAP. Participants received 5 hr of daily 1:1 evidence-based cognitive-linguistically oriented aphasia therapy, supplemented with weekly socially oriented and therapeutic group activities over a 23-day treatment course. Standardized measures of aphasia severity and communicative functioning were obtained at baseline, program entry, program exit, and follow-up. Results were analyzed using a Bayesian latent growth curve model with 2 factors representing (a) the initial level and (b) change over time, respectively, for each outcome measure.

Results: Model parameter estimates showed reliable improvement on all outcome measures between the initial

and final assessments. Improvement during the treatment interval was greater than change observed across the baseline interval, and gains were maintained at follow-up on all measures.

Conclusions: The rationale, clinical processes, and outcomes of a residentially based ICAP have been described. ICAPs differ with respect to treatments delivered, dosing parameters, and outcomes measured. Specifying the defining components of complex interventions, establishing their feasibility, and describing their outcomes are necessary to guide the development of controlled clinical trials.

Key Words: aphasia, outcomes, response to intervention, stroke, effectiveness

In recent years, several intensive comprehensive aphasia programs (ICAPs) have been established in an attempt to improve access to services and to maximize client outcomes. In a recent survey of ICAPs, Rose, Cherney, and Worrall (2013) defined the *service delivery model* as one that provides a minimum of 3 hr of daily treatment over a period of at least 2 weeks and includes individual and group therapy, current technologies, and client/family education to target both impairment and activity/participation levels of language and communicative functioning. The survey revealed considerable variability across programs with respect to the treatment approaches used (e.g., cognitive-linguistic, compensatory, social participation), dosing and

timing attributes (e.g., intensity and duration), and the types of outcomes measured. The report did not provide clinical outcome data from the respective programs surveyed, although one ICAP had previously reported positive outcomes on standardized measures of naming and connected speech in 40 participants with aphasia following 24 hr of weekly therapy for 6 consecutive weeks (Hinckley & Craig, 1998).

In this report, we review the literature addressing the influence of treatment intensity on treatment response in persons with aphasia and discuss additional factors that have served as catalysts for the development of ICAPs. We then describe the rationale, clinical processes, and outcomes of a residentially based intensive aphasia treatment program operated by the VA Pittsburgh Healthcare System (VAPHS).

The influence of treatment intensity on treatment response has received considerable attention in the aphasia literature since Robey's (1998) meta-analysis of 55 quasi-experimental aphasia treatment studies. The analysis included 60 within-subject effects across the 55 studies, 12 of which had examined treatment intensity. The findings

^aVA Pittsburgh Healthcare System, Geriatric Research Education and Clinical Center, Pittsburgh, PA

^bUniversity of Pittsburgh, PA

Correspondence to Patrick J. Doyle: patrick.doyle@va.gov

Editor: Swathi Kiran

Associate Editor: Leora Cherney

Received August 19, 2013

Revision received November 12, 2013

Accepted January 31, 2014

DOI: 10.1044/2014_AJSLP-13-0102

Disclosure: The authors have declared that no competing interests existed at the time of publication.

revealed larger average effect sizes for studies that provided 2 or more hr of treatment per week as compared with those providing less than 2 hr of weekly treatment. Robey concluded that “the more intense the treatment, the greater the change” (p.180).

In a subsequent report, Bhogal, Teasell, and Speechley (2003) examined the relationship between intensity of treatment and treatment outcomes in 10 clinical trials comprising 864 stroke survivors with aphasia. As compared to five trials that reported negative outcomes, the five positive trials provided, on average, more total treatment hours (108 vs. 44) and used a more intensive treatment schedule (8 hr/week vs. 2.5 hr/week) over a shorter treatment interval (11.2 weeks vs. 22.9 weeks). Despite acknowledging some limitations of the primary studies, the authors concluded that “intensive therapy delivered over 2 to 3 months is critical to maximizing aphasia recovery and failure to provide it potentially compromises individual outcomes” (p. 991).

More recently, Cherney and colleagues (Cherney, Patterson, & Raymer, 2011; Cherney, Patterson, Raymer, Frymark, & Schooling, 2008) conducted a series of evidence-based systematic reviews of studies published between 1990 and 2011 that directly compared conditions of higher versus lower doses of aphasia therapy. Eleven studies comprising 195 stroke survivors with aphasia were included in the reviews. Results of these studies were examined with respect to the chronicity of the study sample (acute vs. chronic) and the types of outcome measures obtained (impairment vs. activity/participation). On the basis of the cumulative evidence, Cherney et al. (2011) concluded that no clear advantage emerged across studies for intensive treatment schedules as compared to nonintensive schedules regardless of aphasia chronicity or the types of outcomes measured.

A *Cochrane Collaboration* review (Brady, Kelly, Godwin, & Enderby, 2012) of six randomized controlled clinical trials directly comparing higher versus lower doses of aphasia therapy reported “some indication of the benefits of intensive approaches” (p. 39) as compared to conventional treatment doses. However, the authors qualified these findings, noting a significantly higher dropout rate for participants assigned to intensive treatment conditions and a failure to conduct intention-to-treat analyses in the primary studies.

During the period in which these findings were emerging, changes in reimbursement policy resulted in decreasing lengths of stay for inpatient rehabilitation and shifted the focus of rehabilitation services to outpatient settings (Ottenbacher et al., 2004). The shift to outpatient service delivery and the establishment of per-person Medicare spending limits or therapy caps limited access to care for many community-dwelling persons with aphasia. Demographic, geographic, and transportation factors also represent significant barriers to outpatient care (Arcury, Gesler, et al., 2005; Arcury, Preisser, Gesler, & Powers, 2005; Chapey et al., 2001; Nemet & Bailey, 2000). For example, the Veterans Healthcare Administration (VHA), whose medical centers serve large and frequently rural catchment areas, has experienced high no-show and

clinic cancellation rates (Department of Veterans Affairs Office of the Inspector General, 2008) despite the exemption of their enrollees from Medicare therapy caps. Indeed, a search of the VHA Corporate Data Warehouse (www.virec.research.va.gov/CDW/Overview.htm) revealed that of the 10,371 VHA enrollees with a diagnosis of aphasia in fiscal year 2012, only 3,198 (30%) received speech-language rehabilitation services, with the average amount of treatment totaling 12 hr.

Within this context, the VAPHS established the Program for Intensive Residential Aphasia Treatment and Education (PIRATE), an ICAP that serves community-dwelling veterans and active duty military personnel nationwide. PIRATE participants reside in residential housing provided by the VAPHS for the duration of the 4-week program. During this time, they receive 5 hr of daily 1:1 evidenced-based cognitive-linguistically oriented aphasia therapy, supplemented with weekly socially oriented and therapeutic group activities provided by licensed speech-language pathologists (SLPs) experienced in aphasia assessment and intervention. A total of six program sessions are offered each year with three participants per session, permitting an annual program capacity of 18 participants.

Clinical Processes

Program referral and admission. Referrals to the program are accepted from VHA medical centers and Department of Defense hospitals nationwide. All candidates referred to PIRATE are required to complete a program application that requests family, educational, and employment history as well as information regarding cause of aphasia, date of onset, and prior treatment history. Also completed at the time of application is a detailed medical history form that includes an inventory of clinical symptoms across organ systems (review of systems), current medications, and cardiovascular risk factors. These documents and the electronic medical record are reviewed and discussed by the PIRATE caseworker, SLPs, and physician. Subsequent phone interviews are conducted by team members with the candidate, family or caregivers, referring physician, and local SLP to clarify and further explore information provided in the application materials and medical record. Specific areas of inquiry include mood and behavioral status, level of independence in activities of daily living (ADLs), including medication management, aphasia treatment history, predicted tolerance for intensive treatment, motivation to participate in treatment, family support and involvement in the plan of care, and disqualifying comorbidities (e.g., recent or uncontrolled seizures, uncontrolled hypertension, and cardiac arrhythmias).

Candidates who are determined to be of stable medical and mental health status, independent in ADLs, and able to tolerate an intensive treatment schedule are then scheduled for a comprehensive speech-language evaluation which is the final data source used to determine program admission.

Speech-language evaluation. In addition to serving as a factor for program admission, the initial speech-language assessment is designed to identify the nature and severity of

the candidate's communication disorder and to describe the cognitive and linguistic processing deficits underlying their communication difficulties. Two primary assessment instruments are used for these purposes: (a) the Comprehensive Aphasia Test (CAT; Swinburn, Porter, & Howard, 2004), which provides a measure of overall aphasia severity as well as an analysis of the nature of the language impairment in relation to current psycholinguistic theories of language processing (Howard, Swinburn, & Porter, 2010), and (b) a connected speech sample (Nicholas & Brookshire, 1993), which permits an analysis of the morphological, lexical, and syntactic elements of the candidate's spoken language. On the basis of these findings, selected supplemental tests are administered at the discretion of the examining SLP. These include, but are not limited to, the Psycholinguistic Assessments of Language Processing in Aphasia (PALPA; Kay, Lesser, & Coltheart, 1992), Pyramids and Palm Trees (Howard & Patterson, 1992), the Northwestern Assessment of Verbs and Sentences (NAVS; Cho-Reyes & Thompson, 2012), and the Philadelphia Naming Test (PNT; Roach, Schwartz, & Martin, 1996). These supplemental assessments are administered to provide additional information regarding the nature and locus of linguistic processing deficits identified by the CAT and to guide the development of individualized treatment plans.

For example, selected subtests of the PALPA and error analysis of PNT items can be used to identify impaired lexical-semantic and/or phonological processes within models of word production (e.g., Schwartz, Dell, Martin, Gahl, & Sobel, 2006). Identifying which psycholinguistic processes are impaired may then be used to focus word-finding treatments. Similarly, performance on the NAVS and analysis of the connected speech sample may be used to identify specific sentence-level impairments, for example, whether such impairments are specific to syntactically complex sentences or extend to simple sentences or verb-argument structures. This information can then guide the choice of sentence-level treatments.

On completion of the initial evaluation, the findings are documented in the electronic medical record and reviewed with candidates and their caregivers. Decisions regarding program admission are provided at this time, along with a rationale for any decision not to admit a particular candidate.

Candidates who are offered program admission participate in a separate treatment planning meeting with the PIRATE team to discuss short-term treatment goals, potential psycholinguistically motivated treatment options, personal and environmental factors affecting their communicative functioning, and compensatory strategies aimed at eliminating barriers and facilitating engagement in valued life activities. These discussions proceed from self-selected goals identified on the initial application and are further formulated on the basis of the results of the initial evaluation and discussions with the candidate and caregiver regarding his or her functional communication needs and valued life activities. The selection of treatment stimuli is also informed by these collaborative discussions and, wherever possible, tailored to the identified goals, interests, and life situation of individual participants.

The current program admission rate is 50% of referrals. The three most common exclusions are (a) unacceptable medical risk as determined by chart review, laboratory test results, and physical exam performed by a VHA physician; (b) inability to perform ADLs independently as determined by clinical exam performed by a VHA-licensed occupational therapist; and (c) the presence of nonlanguage cognitive deficits predicted to limit a candidate's potential to participate in or benefit from treatment as determined by chart review, client/family interviews, neuropsychological tests results (when available), and performance on the cognitive screen of the CAT.

Psycholinguistically oriented treatment approaches. The treatments provided in PIRATE are grounded in current psycholinguistic and cognitive neuropsychological approaches to aphasia therapy. As indicated above, participants' language impairments are described in reference to current models of language processing, and treatments motivated by these models are selected and applied. Whenever possible, treatments supported by empirical studies demonstrating treatment efficacy are selected. Examples include, but are not limited to, semantic feature analysis (Boyle & Coelho, 1995); Verb Network Strengthening Treatment (Edmonds, Nadeau, & Kiran, 2009); Treatment of Underlying Forms (Thompson & Shapiro, 2005); phono-motor treatment (Kendall et al., 2008); and lexical, phonological, and interactive spelling treatments (Beeson, Hirsch, & Rewega, 2002; Beeson, Rising, Kim, & Rapcsak, 2010). In addition to sharing a general theoretical basis in psycholinguistic models of language processing, these treatments all heavily rely on practice and repeated drilling of tasks organized around relatively discrete units of language. Although the particular stimuli used are selected to be personally salient for each client, the focus of treatment is on improving the underlying cognitive-linguistic processes rather than on language performance in a particular social context.

For example, a participant demonstrating an impairment of the lexical-semantic system, as evidenced by patterns of performance on assessment procedures such as selected subtests of the PALPA or the PNT, may be treated using semantic feature analysis (Boyle & Coelho, 1995) which is motivated by the spreading-activation theory of semantic processing (Collins & Loftus, 1975). In this psycholinguistically based treatment approach, word retrieval is targeted by verbal generation of semantic features, thereby accessing conceptual information and activating the semantic network of the target. Similarly, a participant demonstrating impairments in the production of syntactically complex sentences, as evidenced by performance on the NAVS and the connected speech sample, may be treated using Treatment of Underlying Forms. This treatment targets the production of complex sentences, explicitly modeling their abstract grammatical structure to facilitate production (Chomsky, 1986).

Group and socially oriented activities. In addition to highly structured model-driven treatments, group and socially oriented therapy activities that allow for conversational coaching and interaction with both familiar and

unfamiliar partners are provided 4 to 6 hr weekly. All group activities are developed and led by licensed SLPs. Groups typically include three to 10 community-dwelling persons with aphasia. For larger groups, as many as three SLPs may serve as facilitators.

One purpose of socially oriented group activities is to promote generalization of performance gains observed during individual treatment tasks to novel tasks, settings, and communicative partners. A second purpose, grounded in emerging evidence for the efficacy of socially oriented treatment approaches, is to enhance clients' social participation by improving their communication environment or their strategies for overcoming environmental barriers to communication (Elman & Bernstein-Ellis, 1999; Simmons-Mackie, 2001; Worrall, 2000).

For example, a participant's communication goal may be to order food in a restaurant. Although the participant may be receiving individual treatment targeting lexical retrieval of food items, among other categories, in the context of semantic feature analysis (Boyle & Coelho, 1995), and verbal production of verb argument structures within the context of Verb Network Strengthening Treatment (Edmonds et al., 2009), he or she may also engage in small-group treatment activities aimed at ordering food in a restaurant. During such activities, the clinician serves as a facilitator, using conversational coaching techniques or scaffolding techniques to promote effective communication and arranging environmental adaptations to enhance the participant's communicative success (Simmons-Mackie, 2001). Such activities may subsequently be extended to novel settings, such as a cafeteria, and to novel communication partners.

Client education. Educational programming on a variety of topics, including aphasia communication strategies, secondary stroke prevention, aphasia advocacy, and living with aphasia, are developed and presented by licensed SLPs weekly. These 60- to 90-min group presentations are provided to participants and their caregivers (when available) and include customized resources based on each participant's physical, emotional, and communicative needs and geographic location. Resources often include both therapeutic and social resources for recovery, including support groups. The final educational session is conducted with individual participants and their respective caregivers. During this session, the treating SLP discusses the participant's performance in PIRATE, makes recommendations regarding outpatient aphasia therapy services, and provides a home treatment program.

Outcome Measures

Three types of outcomes are collected: (a) performance-based, (b) client-reported, and (c) surrogate-reported assessments of communicative functioning. Performance-based assessments are collected at initial evaluation, program entry, program exit, and follow-up and provide information about impairment-level outcomes. Initial evaluations are conducted on a rolling basis, and the time elapsed between the

initial evaluation and program entry is variable (in the current sample, $M = 3.91$ months, $SD = 3.2$ months). The time between program entry and program exit is less variable ($M = 0.76$ months, $SD = 0.09$ months), given that program-entry testing is typically done on the first 1 to 2 days of a session and exit testing is typically done on the last 1 to 2 days (more detail on the timing of exit testing is provided in the Discussion section). Follow-up testing is typically completed approximately 2 months after program exit ($M = 1.9$ months, $SD = 0.42$ months).

Client and surrogate-reported assessments are obtained at initial evaluation, program entry, and at follow-up (rather than program exit), because they are designed to assess clients' and family members' perceptions of communicative functioning within the context of their daily lives. These assessments provide information about activity and participant-level outcomes. Each assessment is described below.

CAT. In addition to its role as a diagnostic assessment, the CAT serves as the primary performance-based outcome assessment. Specifically, we compute the CAT modality mean T score, which serves as an indicator of overall aphasia severity. This composite score measure is derived by averaging the T scores across the eight modalities of the language battery: (a) comprehension of spoken language, (b) comprehension of written language, (c) repetition, (d) naming, (e) spoken picture description, (f) oral reading, (g) writing, and (h) written picture description (Swinburn et al., 2004).

Early in the program's history, the CAT was given only at the initial evaluation, and the Porch Index of Communicative Ability (PICA; Porch, 2001) overall score was used as the primary performance-based outcome measure. To retain those early cases in our outcome data set, we transformed the PICA scores to the same scale as the CAT modality mean T scores using the following procedures: First, using the published norms, we standardized the PICA scores via equipercentile matching with the standard normal distribution. Second, we transformed the standard scores to T scores ($M = 50$, $SD = 10$). Third, using 36 cases for which we had concurrent PICA and CAT scores from the PIRATE initial evaluation, we regressed the CAT modality T score means onto the PICA T scores. We excluded two cases with absolute standardized residuals greater than 2 and obtained R^2 , slope, and intercept estimates of 0.83, 1.36, and -20.49 , respectively. Finally, we used the coefficient estimates to transform the PICA T scores to the scale of the CAT modality mean T scores. In the results we report below, the CAT scores for 33 individuals, representing 113 data points, were based on PICA scores transformed in this manner. Although the CAT and the PICA differ in their specific theoretical emphases, there is considerable overlap in their conceptual orientation toward aphasia, task selection, and scoring. Thus, we believe that their respective overall scores represent essentially the same construct of overall aphasia severity.

Story Retell Procedure. The Story Retell Procedure (SRP) serves as a second performance-based assessment. The SRP is a standardized language elicitation procedure consisting of four forms that have been demonstrated to

yield equivalent measures of 12 operationally defined productive language variables (Doyle et al., 2000). Each form consists of three short stories that are read aloud by the examiner. After each story, the respondent is instructed to retell the story as completely as possible. A different form was administered to participants at each of the four measurement points. Their retells were audio recorded and scored according to standardized procedures described by McNeil, Doyle, Fossett, Park, and Goda (2001) to obtain an estimate of the informativeness of their connected speech, which we report as our outcome measure.

Aphasia Communication Outcome Measure. The Aphasia Communication Outcome Measure (ACOM; Doyle et al., 2013) is a client-reported and surrogate-reported assessment of communicative functioning. The ACOM included in the present analyses was a 38-item version that included content related to everyday speaking, listening, reading, and writing activities. This particular version was created for use by the PIRATE team on the basis of preliminary data from the initial field trial of the ACOM item pool that was ongoing at the time of the program's inception. All responses are given on a 0-to-3 scale describing the effectiveness with which the communication activity in question is accomplished. For the purposes of our outcomes database, the average item scores are transformed to normalized T scores using the sample of 329 stroke survivors and surrogates who ultimately participated in the initial ACOM field trial as a normative reference group.

Method

Participants

At the time of this report, 73 community-dwelling adults (65 veterans and eight active duty personnel) with aphasia have participated in PIRATE since its inception in January 2009. Participants represent enrollees from 16 different Veterans Integrated Service Networks (VISNs) and 26 states. Of these 73 participants, three had participated in two PIRATE sessions; the present analysis included only the first set of observations for these veterans. In three cases, participants who began a PIRATE session did not complete it. In one case, there was concern for an acute stroke, which turned out to be unfounded, and this veteran provided follow-up outcome data and returned for a second session. In the second case, the veteran decided to end his participation early for personal reasons. In the third case, the veteran experienced multiple falls during his first week and was determined to be unsafe for independent living. Of the 73 veterans included in this report, seven were less than 6 months postonset of aphasia at the time of initial evaluation, and two were less than 6 months postonset at program entry. Additional descriptive data are presented in Table 1.

Clinical Outcomes: Analysis and Results

Analysis model. We analyzed the outcome data using a Bayesian latent growth curve model (Zhang, Hamagami, Wang, Nesselroade, & Grimm, 2007). Latent growth curve modeling serves the same purpose in this context as repeated-

Table 1. Demographic and clinical descriptive data for Program for Intensive Residential Aphasia Treatment and Education (PIRATE) participants to date.

Variable	Measure
Gender (% male)	92
Age at program entry in years (<i>M/SD</i>)	55/14
Years of education (<i>M/SD</i>)	14/2.5
Race (%)	
White	80
African American	15
American Indian	1
Pacific Islander	1
Asian	1
Ethnicity: Hispanic or Latino (%)	3
Months post onset at initial evaluation	18/1–227
(<i>Mdn/min–max</i>)	
Months elapsed between initial evaluation and program entry (<i>M/SD</i>)	3.9/3.2
Etiology of aphasia (%)	
Left hemisphere stroke	88
Bilateral stroke	3
Closed head injury	5
Penetrating head injury	3
Herpes encephalitis	1
Motor speech diagnosis (%)	
No motor speech disorder	64
Apraxia of speech	23
Dysarthria	8
Apraxia of speech and dysarthria	3

Note. *N* = 73. min = minimum; max = maximum.

measures analysis of variance (ANOVA) but is associated with the factor analysis tradition and is less exclusively focused on differences between groups (as compared to differences between individuals) than is repeated-measures ANOVA (Voelkle, 2007). Also, latent growth curve modeling can be viewed as a general approach that includes repeated-measures ANOVA as a special case (Voelkle, 2007).

In the present latent growth curve model the three or four observations for each outcome measure were specified as indicators of two latent (unobserved) factors, one representing the initial score level and the other representing change over time. The level-factor coefficients (loadings) were fixed at 1 for all time points. The change-factor coefficients were fixed at 0 and 1 for the first and last observations (baseline and follow-up), respectively, and the coefficients for the intermediate observations (program entry and program exit) were freely estimated. These specifications permitted modeling of nonlinear patterns of change over time. The latent growth factors were permitted to covary within and across the four outcome measures. The residual variances for each observed variable were constrained to be independent and equal across time for each outcome assessment. These residual variance terms explicitly account for measurement error in the observations.

With the model specified as described above, each observed data point is viewed essentially as a function of four quantities: (a) the estimated initial level, which is added to (b) the estimated overall amount of change between the first

and last observations, multiplied by (c) a change coefficient associated with each time point and (d) measurement error. We chose this model because it allowed us to evaluate all four outcome assessments in a single analysis, test for differences in performance at each time point, and estimate correlations between initial level and degree of change both within and across assessments.

We estimated the latent growth curve model using Bayesian methods because they provide a basis for inference that is more consistent with the requirements of clinical program evaluation than traditional frequentist methods based on null-hypothesis significance testing. Whereas traditional frequentist p values and confidence intervals provide information only about the probability of the data given the null hypothesis, Bayesian methods provide information about the hypotheses of interest given the data and specified prior assumptions (Gill, 2009). This information is provided in the form of a distribution for each model parameter of interest, called the *posterior distribution*, to distinguish it from the *prior distribution*, which represents one's knowledge before observing any of the current data. Also, the interpretation of frequentist p values depends on the notion that the observations could, in principle, be independently replicated a large number of times under exactly the same conditions. This dependence on frequency counts as a basis for probabilistic inference leads to well-known concerns about multiple comparisons and the inflation of the Type I error rate (Gelman, Hill, & Yajima, 2012; Kruschke, 2010). In contrast, the Bayesian view of probability is that it represents a subjective degree of belief. Bayesian inference proceeds from prior assumptions that are updated with observed data to yield posterior distributions for each quantity of interest. These posteriors then become the priors for subsequent analyses when additional observations are made. This system obviates concerns about the inflation of the Type I error rate arising from multiple analyses of overlapping data sets. It also lends itself well to clinical program evaluation where routine updating of prior knowledge with new data is desirable and appropriate.

Missing data. Sixty-one participants (80%) had one or more missing observations, and 21.3% of observations were missing overall. Sixty-three participants (83%) provided complete data on at least one outcome measure, and 46 participants (61%) provided complete data on two or more measures. All participants provided at least one pre-PIRATE observation on all four outcome measures, and 68 participants (93%) had at least one posttreatment observation. Three participants who lacked posttreatment observations were recent PIRATE participants for whom posttreatment data were not available at the time of these analyses. The two other participants who lacked posttreatment data were veterans described above who did not complete PIRATE.

A plurality of the missing observations (48% of the missing data points) occurred at follow-up across the four assessments, and we attribute the missingness in these cases at least in part to the difficulties inherent in collecting data from individuals spread over a large geographic area. In many cases, veterans are understandably unable or unwilling

to return to Pittsburgh for follow-up testing, and we have only relatively recently begun to assess PIRATE participants using telehealth methods. As a rough test of this assumption, we coded each participant as to whether the follow-up CAT score was missing and whether the veteran was served by our local VISN (VISN 4). The correlation between these two variables was $-.35$, $p < .01$, indicating that veterans from within VISN 4 were indeed less likely to have missing data at follow-up. At the same time, the VISN 4 indicator variable was unrelated to the estimated level (initial performance) or shape (change) scores. A second large portion of missing observations (an additional 26% of missing data points) occurred for the SRP at all four time points because we began using it as an outcome measure only in the second year of the program. The remaining 26% of missing data points, for which missingness is more difficult to summarize, comprised only 5.5% of the total observations.

In estimating the latent growth curve model, missing data were handled with multiple imputation (Asparouhov & Muthén, 2010; Graham, 2009). We conducted the initial analysis treating all missing observations as missing at random,¹ and we used the proposed model to estimate values for the missing data. Five data sets with imputed values for all missing observations were created; each of these was analyzed; and the final parameter estimates, posterior standard deviations, and credible intervals were taken as the average across the five analyses. Multiple imputation is generally preferred to other methods of dealing with missing data, such as listwise deletion or mean substitution, because it typically introduces less bias into model parameter estimates and more faithfully reflects the increased uncertainty associated with missing observations (Graham, 2009).

Results

The latent growth curve model was estimated in Mplus version 7 (Muthén & Muthén, 2012) using the Bayes estimator with default uninformative priors.² We confirmed Markov chain Monte Carlo (MCMC) convergence using the Gelman–Rubin potential scale reduction factor (Gelman & Rubin, 1992) and inspection of autocorrelations and trace plots for each chain. Details are provided in the Appendix. Final posterior distributions were estimated on the basis of 50,000 MCMC samples from two independent chains after discarding a 50,000-sample burn-in period.

Posterior predictive checking using the likelihood ratio χ^2 statistic (Gelman, Meng, & Stern, 1996) indicated tenable fit of the latent growth curve model to the data for both

¹In the context of missing data analysis, *missing completely at random* means that the cases with missing observations can be considered a random subset of all cases. *Missing at random* means that the missing data are missing completely at random conditional on the observed data.

²Uninformative priors make the weakest possible assumptions and give the current data maximal influence over the resulting posterior distributions and parameter estimates. Bayesian analysis with uninformative priors typically yields results that are very similar to frequentist methods.

the original data set, $p = .073$, and the imputed sets, average $p = .079$. The model-estimated means and 95% credible intervals³ (CIs) for each outcome measure at each observation are provided in Figure 1. For each outcome measure, the mean and variance of the latent change factor were greater than 0, with 95% CIs excluding 0, indicating first that there was, on average, significant positive change between baseline and follow-up on all measures and, second, that participants differed from each other in the degree of change. Distributions of individual change score estimates for each assessment are presented in Figure 2. These estimates are plausible values computed for each of the five multiple imputation data sets (Von Davier, Gonzalez, & Mislevy, 2009), which represent a likely distribution of change scores for each individual. Rather than averaging these plausible values for each participant, we plotted each value separately in the histograms in order to accurately represent the uncertainty in the individual change score estimates. Figure 2 shows that despite the strong average trend for scores to increase over time, in a minority of cases there was likely no change or even decline. At the same time, there was reliable improvement on at least one outcome in 96% of cases, on two or more outcomes in 72% of cases, and on all four outcomes in 24% of cases.

One of the advantages of MCMC methods for Bayesian model estimation is that they conveniently support inferences about arbitrary functions of model parameters. We used functions of the latent factor means and change coefficients to answer specific questions about the patterns of change over time in terms of the score units for each assessment. These results are summarized in Table 2 and described below.

First, we asked whether there was change during the baseline interval—that is, between the initial evaluation and program entry. There was positive change on all four measures, and in each case the 95% CI excluded 0. Thus, we concluded that there was at least a 95% probability that, on average, scores improved on these assessments during the baseline interval. Second, we asked whether there was change during the treatment interval—that is, between program entry and program exit for the CAT and SRP and between program entry and follow-up for the ACOM and surrogate ACOM (SUR-ACOM). In each case, the change score was positive with a 95% CI that excluded 0.

Third, we asked whether the change observed during the treatment interval was greater than the change observed during the baseline interval. The difference was positive and reliable for the CAT and SUR-ACOM, but the 95% CIs for the SRP and the ACOM included 0.

We also computed for each assessment the rate of change (per month) observed during baseline and treatment and compared them. It is important to note that for all four assessments the rate of change observed during treatment

was reliably greater than the rate of change observed during baseline.

Finally, for the CAT and the SRP we asked whether there was significant change between program exit and the follow-up assessment. For the CAT, the estimated score at follow-up was 0.30 T score units higher than at program exit, 95% CI [0.04, 0.58]. On the SRP, the follow-up score was 1.34% information units lower than at program exit, 95% CI [-2.36, -0.50]. In both cases, the change score was significantly different from 0.

In addition to drawing inferences about change over time, we inspected the matrix of correlations between the level and change factors for the four outcome assessments. The full correlation matrix is presented in Table 3. As expected, the level factors for all four assessments were positively correlated, but with wide credible intervals in some cases. The correlations between the change factors were less consistent and had even wider credible intervals, though with positive point estimates in most cases. Also, in general, initial performance was positively correlated with change across the measures, though not always reliably so. The strongest correlations were between the CAT and SRP, and the weakest generally involved the SUR-ACOM.

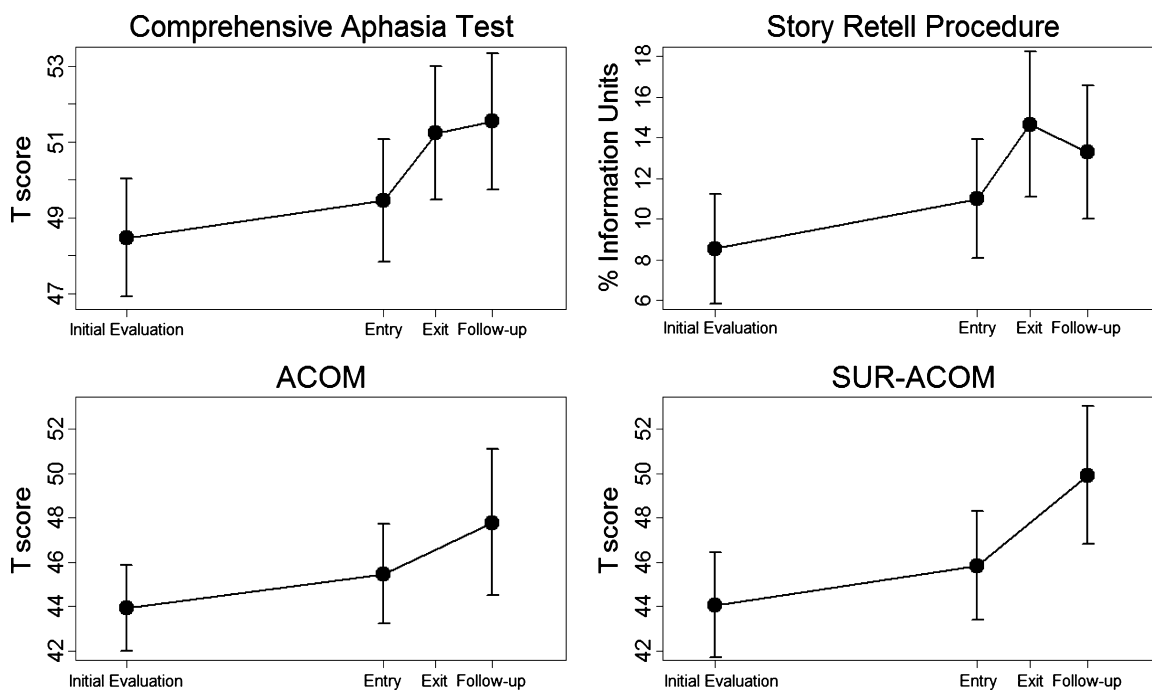
Discussion

An increasing number of ICAPs are being established in both the United States and abroad. These programs differ with respect to the treatments delivered, dosing parameters, and outcomes measured. The purpose of this article was to describe the rationale, clinical processes, and outcomes of a residentially based intensive aphasia treatment program operated by the VAPHS.

We found that a selected group of veterans and active duty military personnel improved on four standardized measures indexing both impairment level and activity/participation level communication outcomes following the 23-day intensive treatment course described above. Specifically, using a Bayesian latent growth model we found positive reliable group change during the baseline period, followed by positive change during the treatment period that was reliably greater in both its magnitude and rate of change. These findings are consistent with those of Code, Torney, Gildea-Howardine, and Willmes (2010) and Rodriguez et al. (2013), who observed improved performance on both impairment-level and activity-/participation-level outcomes in adults with chronic aphasia after intensive, comprehensive aphasia treatment. With respect to the change observed during the baseline period (M months elapsed = 3.9, SD = 3.2 months), it should be noted that a number of the present participants were receiving nonintensive outpatient treatment at their local VA medical centers during this interval and that seven of these participants were also within 6 months postonset of their aphasia. Together, these factors may have accounted for the observed changes during baseline. These findings are consistent with those of Hinckley and Craig (1998), who used a similar within-subject group design and observed greater change on measures of confrontation naming

³Credible intervals are the Bayesian analogue to frequentist confidence intervals. A 95% credible interval is the range that provides a 95% probability of covering the parameter in question, given the data and prior assumptions.

Figure 1. Outcome score estimates by assessment point. The distance between points on the x-axis reflects the average relative time elapsed between assessments. The error bars indicate Bayesian 95% credible intervals. ACOM = Aphasia Communication Outcome Measure; SUR-ACOM = surrogate ACOM.



and the informativeness of connected speech following an intensive residentially based treatment program (23 hr/week for 6 weeks) as compared to a nonintensive outpatient treatment schedule (< 3 hr/week for 6 weeks) in a sample of 15 community-dwelling adults with chronic aphasia.

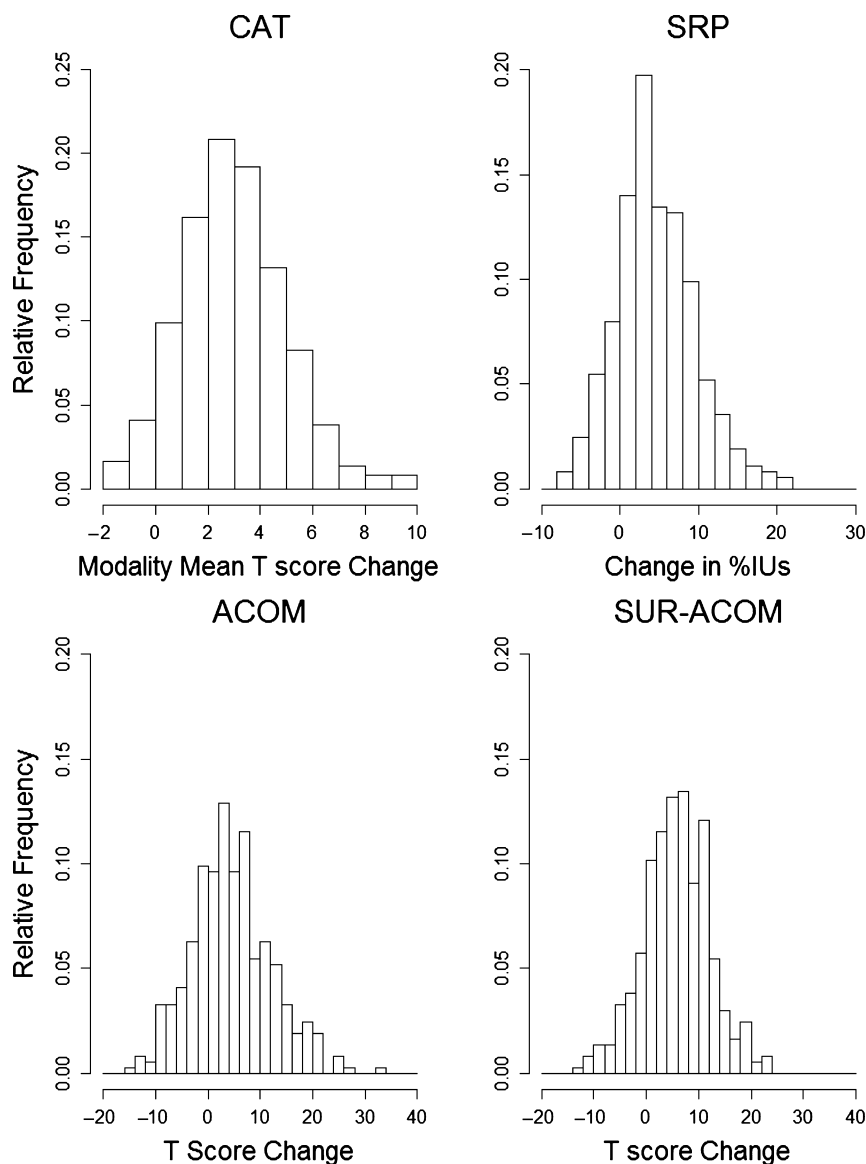
We also should noted that although participants who improved over the course of PIRATE tended to do so on multiple outcomes, not all participants who completed the program improved. Figure 2 shows that, for a small minority of participants, no change or a decline in performance was observed for all outcomes—that is, despite our efforts to select community-dwelling candidates considered to have a positive treatment prognosis based on their individual medical, mental health, and cognitive and social support status, we did not observe universal improvement on our selected outcome measures. This finding is consistent with those of Code et al. (2010), Hinckley and Craig (1998), and Rodriguez et al. (2013), who reported large individual differences in response to intensive treatment in independent samples of adults with chronic aphasia.

Although it was beyond the scope of this project to identify predictors of treatment success, readers should note in Table 3 that aphasia severity as measured by the CAT modality mean T score at initial evaluation was moderately correlated (.39–.58) with change over time on our outcome measures. Thus, in this sample of community-dwelling adults with chronic aphasia, aphasia severity accounted for 15%–30% of the variance in individual outcomes. This finding is

consistent with those of Code et al. (2010), who reported that their more severely impaired participants improved less, but it is inconsistent with those of Persad, Wozniak, and Kostopoulos (2013), who reported negative and null associations between aphasia severity and intensive treatment outcomes in a sample of 125 adults with chronic aphasia. Our findings are also inconsistent with those of Dickey and Yoo (2010), who reported a null association between aphasia severity and treatment outcomes in their meta-analysis of 30 participants who underwent linguistically motivated sentence production training. Further well-designed prospective research trials will be required to identify accurate predictors of intensive treatment outcomes.

Although the findings reported herein are promising, they must be interpreted with caution and are subject to the following limitations. First, in this article, we have described an ongoing retrospective evaluation of the feasibility and outcomes of a novel service delivery model intended to increase access to evidence-based language rehabilitation services and to improve the outcomes of community-dwelling adults with aphasia. As such, the within-subject time series analysis of the data was sufficient for the intended observational purposes of this project, but it does not permit strong inferences regarding the relationship between the applied intervention and the observed outcomes. Thus, we can confidently conclude that our participants improved and that their improvement was greater in magnitude and rate during treatment as compared to at baseline. Nevertheless, the lack

Figure 2. Histograms of plausible values for estimated factor scores representing the change between initial evaluation and follow-up for each participant. For each distribution, the plotted values greater than 0 represent the estimated proportion of participants whose scores improved on that measure. CAT = Comprehensive Aphasia Test; SRP = Story Retell Procedure; IUs = information units.



of an appropriate control group precludes us from inferring that our treatments caused the improvement. For example, we cannot rule out the potential influence of known extraneous variables, such as the communal living situation in which our participants resided during the 24-day residential program, or other unknown extraneous variables or their interaction with the active intervention program.

Second, all participants received individually tailored treatment protocols that frequently included multiple psycholinguistically oriented treatments delivered in parallel or sequentially during the course of treatment as indicated, as well as more socially oriented group activities. Thus, only the

intensity, duration, and general treatment approach were held constant across participants; there was no attempt to deliver a predetermined treatment protocol to all program participants. Therefore, we cannot assert with confidence which components of this complex intervention may have been the active agents of the observed changes (see Baker, 2012; Cherney, 2012; and Warren, Fey, & Yoder, 2007, for further discussion of this issue).

Third, our sample was selected only from the population of VHA-enrolled veterans and active duty service personnel who were referred to the program, and program admission was determined by team consensus that was based

Table 2. Comparisons regarding change over time and rate of change per month on each outcome assessment.

Assessment	Comparison	2.5th percentile	Point estimate	97.5th percentile
CAT (T score units)	Change during baseline	0.68	0.97	1.29
	Change during treatment	1.40	1.78	2.19
	Treatment change – baseline change	0.34	0.80	1.30
	Rate of change during baseline	0.17	0.25	0.33
	Rate of change during treatment	1.83	2.33	2.87
	Treatment change rate – baseline change rate	1.59	2.08	2.62
	Change during follow-up	0.04	0.30	0.58
SRP (% IUs)	Change during baseline	1.43	2.43	3.61
	Change during treatment	2.34	3.66	5.12
	Treatment change – baseline change	-0.18	1.19	2.78 ^a
	Rate of change during baseline	0.37	0.62	0.92
	Rate of change during treatment	3.06	4.79	6.71
	Treatment change rate – baseline change rate	2.53	4.16	6.04
	Change during follow-up	-2.36	-1.34	-0.50
ACOM (T score units)	Change during baseline	0.28	1.47	3.07
	Change during treatment	0.55	2.27	4.43
	Treatment change – baseline change	-0.87	0.69	2.99 ^a
	Rate of change during baseline	0.07	0.38	0.79
	Rate of change during treatment	0.30	1.25	2.43
	Treatment change rate – baseline change rate	0.09	0.83	2.02
	Change during follow-up	-2.36	-1.34	-0.50
SUR-ACOM (T score units)	Change during baseline	0.55	1.75	3.19
	Change during treatment	2.21	4.04	6.08
	Treatment change – baseline change	0.13	2.22	4.85
	Rate of change during baseline	0.14	0.45	0.82
	Rate of change during treatment	1.22	2.22	3.34
	Treatment change rate – baseline change rate	0.75	1.75	2.99

Note. CAT = Comprehensive Aphasia Test; SRP = Story Retell Procedure; IUs = information units; ACOM = Aphasia Communication Outcome Measure; SUR-ACOM = surrogate ACOM. The data presented are the 2.5th percentile, median, and 97.5th percentile of the posterior distribution for each comparison, averaged across analyses of five data sets with imputed values for missing observations.

^a95% credible interval includes 0.

on a number of quantitative and qualitative variables. Thus, we cannot rule out the influence of selection bias on the observed results.

Fourth, despite our efforts to collect complete outcome data from all program participants, a nonnegligible proportion of observations (21%) were missing in the present analysis, and this may have biased the results. We took steps to investigate and mitigate the potential impact of the missing data, including identifying potential causes of missingness that were unrelated to program outcome (loss at follow-up due

to distance from Pittsburgh, and the late addition of the SRP to the program outcome battery) and using multiple-imputation methods to more accurately characterize the uncertainty in the model parameter estimates. Nevertheless, the missing data are a limitation of this report.

Fifth, certain program components were modified during the course of its development in response to practical constraints and organizational opportunities and priorities. Two such modifications involved changes to the testing environment in which the reported outcomes were obtained

Table 3. Matrix of estimated correlations between the latent growth factors (initial level and change over time) for each of the four outcome assessments.

Variable	1	2	3	4	5	6	7	8
1. CAT level	—							
2. CAT change	.39 (.10, .62)	—						
3. SRP level	.78 (.64, .87)	.17 (-.14, .45)	—					
4. SRP change	.42 (.12, .65)	.70 (.44, .87)	.28 (-.05, .57)	—				
5. ACOM level	.41 (.08, .67)	.52 (.17, .77)	.33 (-.01, .61)	.32 (-.07, .64)	—			
6. ACOM change	.58 (.27, .80)	.48 (.10, .75)	.43 (.08, .70)	.47 (.08, .76)	.41 (-.05, .78)	—		
7. SUR-ACOM level	.33 (.04, .56)	.08 (-.24, .38)	.35 (.06, .59)	.05 (-.28, .37)	.50 (.16, .74)	.27 (-.11, .59)	—	
8. SUR-ACOM change	.36 (.002, .66)	-.14 (-.51, .27)	.24 (-.13, .57)	.02 (-.39, .41)	.16 (-.28, .56)	.38 (-.06, .73)	.12 (-.27, .53)	—

Note. Credible intervals (95%) are presented in parentheses, and estimates with credible intervals excluding 0 are in boldface type.

and could potentially serve as additional alternative explanations for certain aspects of the results.

One modification involved moving the CAT and SRP program-exit assessments from the last day of the program, when many participants were reunited with family members and frequently had air travel arrangements, to the day prior. The original arrangement was considered a limitation by program participants and family members, and our clinical providers expressed concern that it was contributing to lower performance on the outcome assessments. Indeed, separate analyses of participants tested on the last day ($n = 38$) as compared to the penultimate day ($n = 35$) revealed changes from exit to follow-up on the CAT only in the former and not the latter cohort. This raises the possibility that the improvement at follow-up reported for the total sample may represent an artifact of the different assessment conditions and that CAT performance was actually depressed in participants whose program-exit evaluations were conducted on the same day as their travel arrangements.

Another program modification involving the testing environment concerned the methods of assessment used for initial evaluations and program follow-up. Because of both an increasing number of referrals from geographically remote VA medical centers and VHA priorities, in February 2012 we began conducting initial evaluations and follow-up assessments via *clinical video telehealth*, a modality in which standard telehealth equipment and a high-bandwidth connection provide a secure, real-time audio and video connection between two clinical sites. Prior to this time, follow-up CAT and SRP assessments were administered and scored face to face by the local (referring) SLP. Current studies that have included direct comparisons of the SRP (Georgeadis, Brennan, Barker, & Baron, 2004) and the CAT (Winans-Mitrik et al., 2013) administered via telehealth and face to face have reported that these and other aphasia assessments (Hill, Theodoros, Russell, Ward, & Wootton, 2009) yielded comparable results across testing conditions when delivered by trained providers experienced with the test instrument. In the current data, separate analysis of the SRP data in the 17 participants who had follow-up assessment done by a PIRATE clinician and the 11 who had a follow-up SRP administered and scored by a referring clinician revealed a smaller decline in the former group (1.5% information units [IUs] vs. 4.5% IUs), though in both cases the 95% CI included 0. Thus, although it is not possible to assert with confidence, the decrease in performance on the SRP observed from program exit to program follow-up may be explained in part by differences in the reliability between local SLPs who administered and scored the SRPs online versus SRPs obtained via clinical video telehealth that were recorded and scored offline by PIRATE SLPs who were more experienced with the SRP administration and scoring procedures.

In this article, we have described the rationale, clinical processes, and clinical outcomes of a residentially based ICAP. ICAPs differ with respect to treatments delivered, dosing parameters, and outcomes measured. Specifying the defining components of complex interventions, establishing

their feasibility, and describing their outcomes are necessary to guide the development of controlled clinical trials. Future research should be aimed at establishing what client factors contribute to treatment success, what components of complex intervention are active agents of the observed changes, and the treatment dose necessary to maximize both impairment and activity-/participation-level communication outcomes in community-dwelling adults with aphasia.

Acknowledgments

PIRATE is a clinical demonstration project supported by VA Pittsburgh's Geriatric Research Education and Clinical Center. We wish to acknowledge the invaluable contributions of Shannon Austermann Hula, Carol Dolbee, Malcolm R. McNeil, Michelle Rossi, and Mary Sullivan, as well as both past and current PIRATE speech pathologists Michael Biel, Geoffrey Fredericks, Beth Friedman, Christine Matthews, Rebecca Owens, and Cheralyn Ranjan.

References

- Arcury, T. A., Gesler, W. M., Preisser, J. S., Sherman, J., Spencer, J., & Perin, J. (2005). The effects of geography and spatial behavior on health care utilization among the residents of a rural region. *Health Services Research, 40*, 135–156.
- Arcury, T. A., Preisser, J. S., Gesler, W. M., & Powers, J. M. (2005). Access to transportation and health care utilization in a rural region. *The Journal of Rural Health, 21*, 31–38.
- Asparouhov, T., & Muthén, B. O. (2010). *Multiple imputation with Mplus*. Retrieved from www.statmodel.com/download/Imputations7.pdf
- Baker, E. (2012). Optimal intervention intensity. *International Journal of Speech-Language Pathology, 14*, 401–409.
- Beeson, P. M., Hirsch, F. M., & Rewega, M. A. (2002). Successful single-word writing treatment: Experimental analyses of four cases. *Aphasiology, 16*, 473–491.
- Beeson, P. M., Rising, K., Kim, E. S., & Rapsak, S. Z. (2010). A treatment sequence for phonological alexia/agraphia. *Journal of Speech, Language, and Hearing Research, 53*, 450–468.
- Bhagal, S. K., Teasell, R., & Speechley, M. (2003). Intensity of aphasia therapy: Impact on recovery. *Stroke, 34*, 987–993.
- Boyle, M., & Coelho, C. A. (1995). Application of semantic feature analysis as a treatment for aphasic dysnomia. *American Journal of Speech-Language Pathology, 4*, 94–98.
- Brady, M. C., Kelly, H., Godwin, J., & Enderby, P. (2012). Speech and language therapy for aphasia following stroke. *Cochrane Database of Systematic Reviews, 2012*(5), Article CD000425.
- Chapey, R., Duchan, J. F., Elman, R. J., Garcia, L. J., Kagan, A., Lyon, J. G., & Simmons-Mackie, N. (2001). Life participation approach to aphasia: A statement of values for the future. In R. Chapey (Ed.), *Language intervention strategies in aphasia and related neurogenic communication disorders* (4th ed., pp. 235–245). Baltimore, MD: Lippincott Williams & Wilkins.
- Cherney, L. R. (2012). Aphasia treatment: Intensity, dose parameters, and script training. *International Journal of Speech-Language Pathology, 14*, 424–431.
- Cherney, L. R., Patterson, J. P., & Raymer, A. M. (2011). Intensity of aphasia therapy: Evidence and efficacy. *Current Neurology & Neuroscience Reports, 11*, 560–569.
- Cherney, L. R., Patterson, J. P., Raymer, A., Frymark, T., & Schooling, T. (2008). Evidence-based systematic review: Effects

- of intensity of treatment and constraint-induced language therapy for individuals with stroke-induced aphasia. *Journal of Speech, Language, and Hearing Research*, 51, 1282–1299.
- Cho-Reyes, S., & Thompson, C. K.** (2012). Verb and sentence production and comprehension in aphasia: Northwestern Assessment of Verbs and Sentences (NAVS). *Aphasiology*, 26, 1250–1277.
- Chomsky, N.** (1986). *Knowledge of language: Its nature, origins, and use*. Westport, CT: Greenwood Press.
- Code, C., Torney, A., Gildea-Howardine, E., & Willmes, K.** (2010). Outcome of a one-month therapy intensive for chronic aphasia: Variable individual responses. *Seminars in Speech and Language*, 31, 21–33.
- Collins, A. M., & Loftus, E. F.** (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82, 407–428.
- Department of Veterans Affairs Office of the Inspector General.** (2008). *Audit of Veterans Health Administration's efforts to reduce unused outpatient appointments* (Report No. 08-00879-36). Washington, DC: Author.
- Dickey, M. W., & Yoo, H.** (2010). Predicting outcomes for linguistically specific sentence treatment protocols. *Aphasiology*, 24, 787–801.
- Doyle, P. J., Hula, W. D., Hula, S. N. A., Stone, C. A., Wambaugh, J. L., Ross, K. B., & Schumacher, J. G.** (2013). Self- and surrogate-reported communication functioning in aphasia. *Quality of Life Research*, 22, 957–967.
- Doyle, P. J., McNeil, M. R., Park, G. H., Goda, A. J., Rubenstein, E., Spencer, K., ... Szwarc, L.** (2000). Linguistic validation of four parallel forms of a story retell procedure. *Aphasiology*, 14, 537–549.
- Edmonds, L. A., Nadeau, S. E., & Kiran, S.** (2009). Effects of Verb Network Strengthening Treatment (VNeST) on lexical retrieval of content words in sentences in persons with aphasia. *Aphasiology*, 23, 402–424.
- Elman, R. J., & Bernstein-Ellis, E.** (1999). The efficacy of group communication treatment in adults with chronic aphasia. *Journal of Speech, Language, and Hearing Research*, 42, 411–419.
- Gelman, A., Hill, J., & Yajima, M.** (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness*, 5, 189–211.
- Gelman, A., Meng, X. L., & Stern, H.** (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, 6, 733–760.
- Gelman, A., & Rubin, D. B.** (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457–511.
- Georgeadis, A., Brennan, D., Barker, L., & Baron, C.** (2004). Telerehabilitation and its effect on story retelling by adults with neurogenic communication disorders. *Aphasiology*, 18, 639–652.
- Gill, J.** (2009). *Bayesian methods: A social and behavioral sciences approach* (2nd ed.). London, UK: CRC Press.
- Graham, J. W.** (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, 60, 549–576.
- Hill, A. J., Theodoros, D. G., Russell, T. G., Ward, E. C., & Wootton, R.** (2009). The effects of aphasia severity on the ability to assess language disorders via telerehabilitation. *Aphasiology*, 23, 627–642.
- Hinckley, J. J., & Craig, H. K.** (1998). Influence of rate of treatment on the naming abilities of adults with chronic aphasia. *Aphasiology*, 12, 989–1006.
- Howard, D., & Patterson, K.** (1992). *Pyramids and palm trees: A test of semantic access from pictures and words*. Windsor, UK: Thames Valley Test Company.
- Howard, D., Swinburn, K., & Porter, G.** (2010). Putting the CAT out: What the Comprehensive Aphasia Test has to offer. *Aphasiology*, 24, 56–74.
- Kay, J., Lesser, R., & Coltheart, M.** (1992). *Psycholinguistic assessments of language processing in aphasia*. Hove, UK: Erlbaum.
- Kendall, D. L., Rosenbek, J. C., Heilman, K. M., Conway, T., Klenberg, K., Gonzalez Rothi, L. J., & Nadeau, S. E.** (2008). Phoneme-based rehabilitation of anomia in aphasia. *Brain & Language*, 105, 1–17.
- Kruschke, J. K.** (2010). What to believe: Bayesian methods for data analysis. *Trends in Cognitive Sciences*, 14, 293–300.
- McNeil, M. R., Doyle, P. J., Fossett, T. R. D., Park, G. H., & Goda, A. J.** (2001). Reliability and concurrent validity of the information unit scoring metric for the story retelling procedure. *Aphasiology*, 15, 991–1006.
- Muthén, L. K., & Muthén, B. O.** (2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Authors.
- Nemet, G. F., & Bailey, A. J.** (2000). Distance and health care utilization among the rural elderly. *Social Science & Medicine*, 50, 1197–1208.
- Nicholas, L. E., & Brookshire, R. H.** (1993). A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *Journal of Speech and Hearing Research*, 36, 338–350.
- Ottenbacher, K. J., Smith, P. M., Illig, S. B., Linn, R. T., Ostir, G. V., & Granger, C. V.** (2004). Trends in length of stay, living setting, functional outcome, and mortality following medical rehabilitation. *Journal of the American Medical Association*, 292, 1687–1695.
- Persad, C., Wozniak, L., & Kostopoulos, E.** (2013). Retrospective analysis of outcomes from two intensive comprehensive aphasia programs. *Topics in Stroke Rehabilitation*, 20, 388–397.
- Porch, B. E.** (2001). *Porch Index of Communicative Ability—4th ed.* Albuquerque, NM: PICA Programs.
- Roach, A., Schwartz, M. F., & Martin, N.** (1996). The Philadelphia Naming Test: Scoring and rationale. In M. L. Lemme (Ed.), *Clinical aphasiology* (pp. 121–133). Austin, TX: Pro-Ed.
- Robey, R. R.** (1998). A meta-analysis of clinical outcomes in the treatment of aphasia. *Journal of Speech, Language, and Hearing Research*, 41, 172–187.
- Rodriguez, A. D., Worrall, L., Brown, K., Grohn, B., McKinnon, E., Pearson, C., ... Copland, D. A.** (2013). Aphasia LIFT: Exploratory investigation of an intensive comprehensive aphasia programme. *Aphasiology*, 27, 1339–1361.
- Rose, M., Cherney, L. R., & Worrall, L.** (2013). Intensive comprehensive aphasia programs: An international survey of practice. *Topics in Stroke Rehabilitation*, 20, 379–387.
- Schwartz, M. F., Dell, G. S., Martin, N., Gahl, S., & Sobel, P.** (2006). A case-series test of the interactive two-step model of lexical access: Evidence from picture naming. *Journal of Memory and Language*, 54, 228–264.
- Simmons-Mackie, N.** (2001). Social approaches to aphasia intervention. In R. Chapey (Ed.), *Language intervention strategies in aphasia and related neurogenic communication disorders* (4th ed., pp. 246–267). Baltimore, MD: Lippincott Williams & Wilkins.
- Swinburn, K., Porter, G., & Howard, D.** (2004). *Comprehensive Aphasia Test*. New York, NY: Psychology Press.
- Thompson, C. K., & Shapiro, L.** (2005). Treating agrammatic aphasia within a linguistic framework: Treatment of Underlying Forms. *Aphasiology*, 19, 1021–1036.
- Voelkle, M. C.** (2007). Latent growth curve modeling as an integrative approach to the analysis of change. *Psychology Science*, 49, 375–414.

-
- Von Davier, M., Gonzalez, E., & Mislevy, R. (2009). What are plausible values and why are they useful? *IERI Monograph Series*, 2, 9–36.
- Warren, S. F., Fey, M. E., & Yoder, P. J. (2007). Differential treatment intensity research: A missing link to creating optimally effective communication interventions. *Mental Retardation and Developmental Disabilities Research Reviews*, 13, 70–77.
- Winans-Mitrik, R., Chen, S., Owens, R., Hula, W., Eichhorn, K., Ebrahimi, M., ... Doyle, P. (2013, May). *Tele-rehabilitation solutions for patients with aphasia*. Poster presented at the annual meeting of the Association of Veterans Affairs Speech-Language Pathologists, San Francisco, CA.
- Worrall, L. (2000). A conceptual framework for a functional approach to acquired neurogenic disorders of communication and swallowing. In L. Worrall & C. Frattali (Eds.), *Neurogenic communication disorders: A functional approach* (pp. 3–18). New York, NY: Thieme.
- Zhang, Z., Hamagami, F., Wang, L. L., Nesselroade, J. R., & Grimm, K. J. (2007). Bayesian analysis of longitudinal data using growth curve models. *International Journal of Behavioral Development*, 31, 374–383.

Appendix

MCMC Convergence

In the initial analysis of the original data set and each of the five multiple imputation data sets, we ran two independent MCMC chains for 100,000 iterations. In all cases, the potential scale reduction statistic was lower than 1.01 after 50,000 iterations and had reached 1.001 by 100,000 iterations, suggesting satisfactory convergence. A few model parameters, however, including the variances of the SRP, ACOM, and SUR-ACOM change factors and their covariances, obtained chains with lag-30 autocorrelations in the range of .1 to .25. The trace plots for these chains suggested stable convergence but slow mixing consistent with the elevated autocorrelations, which can lead to underestimation of parameter variability. As a further check of convergence and mixing, we increased the number of MCMC iterations to 200,000 and discarded the first 100,000 as burn-in. The potential scale reduction statistic remained ≤ 1.001 for all parameters. We compared the parameter estimates and posterior standard deviations obtained from this run to the prior run of 100,000 iterations. The estimates differed trivially, by $< 6\%$ of their respective posterior standard deviations, and the posterior standard deviations themselves differed by $\leq 3\%$. For the original data set we also ran a single chain of 1,600,000 iterations, discarded every second iteration due to computer memory limitations, and confirmed that these results produced from the second half of these samples were consistent with those reported above.
