Published in final edited form as:

Aphasiology. 2015 ; 29(5): 570–574. doi:10.1080/02687038.2014.987049.

The case for single-case studies in treatment research— comments on Howard, Best and Nickels “Optimising the design of intervention studies: critiques and ways forward”

Nadine Martin* and Michelene Kalinyak-Fliszar

Department of Communication Sciences and Disorders, Temple University, Philadelphia, PA, USA

Howard, Best, and Nickels (2014) provide an informative state-of-the-art summary of the evolution of single-case treatment research design and some welcome suggestions for designs and statistics to increase the validity of this approach to treatment research. For many aphasiologists, the need for and value of the single-case design is well appreciated, but is challenged by the need for statistical methods and design that allow attribution of language improvement to the treatment. In step with previous advances in design of single-case study research (e.g., Fisher, Kelley, & Lomas, 2003; Howard, 1986; McReynolds & Kearns, 1983; Robey, Schultz, Crawford, & Sinner, 1999; Thompson, 2006, for review), Howard et al. offer a case for considering some methodological adjustments to our current approach to single-case designs for treatment research that could improve their validity and in so doing bolster the acceptance of this approach as an alternative to the medical model of treatment validation.

Historically, patient-oriented research in the communication sciences has faced numerous challenges to its acceptance by mainstream cognitive psychology as a valid approach to the study of the cognitive organisation of language. Levelt, Roelofs, and Meyer (1999), for example, suggested that although it is appropriate to use experimental paradigms developed to study normal language processing for studies of aphasia, we should not expect behaviour of a damaged system to conform to a theory of normal language. This cautionary note side-stepped the challenge of developing a model of language processing that could account for both normal and impaired language behaviours. In fact, cognitive neuropsychological research has played an important role in testing the validity of models of normal language processing by examining their ability to account for impaired language (e.g., Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Howard & Franklin, 1988; Patterson & Shewell, 1987; Rapp & Goldrick, 2000).

Levelt et al. (1999) noted further that a contribution of neuropsychology should be to identify explicit component processes of word production. It is, in fact, this line of research in cognitive neuropsychology, identification of associations and dissociations of components and processes of the language system, which has informed our models of diagnosis and treatment of aphasia used in research and clinical practice. Impairment-based treatments of aphasia have evolved from a focus on communication task-level training (naming,

*Corresponding author. nmartin@temple.edu.

comprehension, repetition) to treatments based on more nuanced descriptions of language impairment that include type of linguistic representations affected (e.g., semantic, phonological, syntactic), processing components affected (access, retrieval, short-term maintenance) and involvement of other cognitive processes (e.g., working memory, executive functioning). Although language impairment profiles are unique to the individual, impairment-based treatments are not necessarily individualised according to a person's language profile. Nonetheless, from these increasingly rich and detailed descriptions of the spared and impaired language abilities of individuals with aphasia, aphasia researchers have developed treatment approaches that can target these impairments more precisely by varying stimulus characteristics (e.g., imageability, word length, semantic, phonological) and stimulus presentation variables (e.g., modality, timing, context).

It is this unique dynamic of matching treatment to impairment profile and goals of the person with aphasia that has been central to the science and “art” of aphasia rehabilitation, but at the same time presents challenges in testing the efficacy of behavioural treatments for aphasia. Randomised clinical trial (RCT) studies that are typically used to test the effectiveness of pharmaceutical treatments are not ideal for testing the efficacy of behavioural treatments in aphasia. Although this issue has been discussed at length elsewhere (e.g., Elman, 2006; Rothi & Barrett, 2006; Whyte, Gordon, & Rothi, 2009), there are several arguments against the use of RCTs that are worth noting because of their relevance to single-case study research. As noted by Rothi and Barrett (2006) and Whyte et al. (2009), the emphasis on RCTs for rehabilitation research minimises the value of single case-studies to establish “proof of concept”. This early stage is theoretically driven and the outcomes differ from later stages of efficacy and effectiveness. Additionally, a single outcome for all participants in a treatment study, as required by an RCT model, is difficult to reconcile with many current treatments designed to target more detailed profiles of language impairment. The impairment and treatment would need to be broadly defined (e.g., semantic vs. phonological treatment for semantic impairment), and this would reduce the clinical and theoretical impact of the treatment, as it would be difficult to know what aspect of the treatment was the source of outcome.

The debate on approaches to assessing efficacy of behavioural treatments for language impairment will continue. Presently, however, much of the research in aphasia rehabilitation uses the single-case or case series model. Despite the wide acceptance of the single-case model, aphasiologists also recognise the need for statistics and designs that can accurately attribute improvements in language behaviour to the treatment. Howard et al. offer some valid criticisms of and suggestions for some current practices in single-case treatment design that could be addressed with some simple changes in application of design and statistical methods. We will discuss three of these suggestions and make some final comments about their recommendation for Weighted Statistics (WEST) to examine treatment effects.

Visual analysis of treatment data

Howard et al. point out the limitations of the visual analysis approach as a means of determining a treatment effect. These shortcomings have been recognised and efforts to address them have been offered by Fisher et al. (2003) and others (e.g., Robey et al., 1999;

Swoboda, Kratochwill, & Levin, 2010). Shewhart (1931) trend lines advocated by Robey et al. (1999) set criterion for a significant effect as two consecutive treatment probes above a line established at 2 *SDs* above the mean of the baseline probe data. Another approach to supplement visual analysis is the Conservative Dual Criterion (CDC) method (Fisher et al., 2003; Swoboda et al., 2010). The CDC method provides a means by which visual inspection of treatment (and follow-up) data can be reliably judged by more than one examiner. A significant effect is determined when a specified number of treatment probes lie above a trend line and a regression line (based on the mean of baseline probe data). In addition, the CDC method has been shown to be more effective in reducing Type I errors than other methods, especially when data are autocorrelated (Fisher et al., 2003).

There are two reasons why we recommend using visual analysis with supplemental methods to establish criteria for evidence of a significant effect. First, this analysis can provide evidence of an effect in cases of a more severe language impairment when treatment gains result in low effect sizes and lack of significant differences between pretreatment and posttreatment measurements. This is important in the overall effort to evaluate whether a treatment has an effect. Second, although the pretreatment and posttreatment measures provide a critical measure of treatment effect and the maintenance of that effect over follow-up measures, “structured” visual analysis of the probe data will inform us about what transpired over the course of treatment. Did improvement on the treatment tasks move in a steady upward trend or was it variable? This information provides insight into trends in the treatment data that individual scores do not.

Effect sizes

When first introduced into the arsenal of measurements and methods to validate treatment efficacy, effects sizes were welcomed enthusiastically and with the expectation of a concrete measure of a treatment’s effect. Howard et al. highlight a number of issues that weaken the usefulness of this measure, some of which have been raised by others (e.g., Beeson & Robey, 2006). They note that the small numbers of treatment items in many studies increases the likelihood of a bias for larger effect sizes. Increasing the number of treatment items may lead to longer treatment sessions or fewer administrations of the treatment per item, and yet, this seems a simple means of minimising that bias.

We have encountered problems with interpretation of effects sizes in three situations. First, much of our treatment research aims to improve repetition and verbal short-term memory. Benchmarks for interpretation of effect sizes (small, medium, large) have been established for treatment of naming and reading disorders (Beeson & Robey, 2006) but not for repetition treatments. Second, it is difficult to interpret the meaning of lower effect sizes when baseline is higher, which is often the case when the language impairment is milder or when a treatment is replicated. Third, when someone’s response to a treatment is modest, effects sizes will be low, and although this may accurately reflect response to treatment, it does not necessarily mean the treatment is ineffective. Severity of impairment is a variable that must be considered in interpreting low effect sizes. A treatment designed to target a particular impairment of language (e.g., repetition of words of low imageability) may be effective if that impairment is moderate, but less so with more severe impairments. In such cases, it is

useful to have statistical measurements of trends of improvement. Additionally, a case series design would be beneficial in dealing with severity and response to treatment, as we would predict that severity would correlate with some measure of improvement that includes both trend and significant levels of effect.

Clinical significance

Howard et al. note that the clinical significance of a treatment is difficult to assess. Although it has been defined as a change that makes a difference in a treatment participant's life (Franklin, Ranklin, Allison, & Gorman, 1997), this depends on what the participant considers to be a significant impact on their life. Like Howard et al., we do not dispute the point that a goal of aphasia rehabilitation is to have a positive impact on the life of the person receiving treatment. However, we add one additional consideration: The onus of impacting the life of someone with aphasia, or their communication abilities in everyday situations, should not rest entirely on the outcomes of impairment-based therapy. If treatment of word retrieval is administered in a context with few opportunities to use improved word retrieval skills in functional communication situations, it seems unrealistic to expect that quality of life automatically would be improved. It would be equally unrealistic to expect that word retrieval skills would necessarily improve if someone is in a supportive environment that promotes quality of life. Ideally, rehabilitation should include tandem administration of impairment-based and life-participation approaches to improve the quality of life for people with aphasia.

Weighted Statistics to compare rate of change across pretreatment and posttreatment phases

Howard et al. offer an alternative to current methods for comparing baseline and posttreatment data, WEST. This group of statistics may be applied under conditions where there is improvement during baseline (WEST Rate of Change), no change in baseline performance (WEST Compare Level of performance) or to determine an overall trend in performance. Although WEST involves both simple and more complex calculations to reduce a set of scores for each item to a single score, this practice could minimise the problem of autocorrelation in treatment data. Additionally, since item scores are multiplied by a weighting that adjusts for the null hypothesis, WEST may be more likely to reveal a true effect of treatment. We suggest the use of visual analysis with trend and regression lines along with the WEST will provide converging evidence of a significant treatment effect. In order to assist the reader in the application of WEST, the authors have wisely included detailed examples of WEST, including weightings for a variety of designs and a procedure to generate the weightings.

In closing, we applaud Howard et al. assessment of current methods to evaluate the efficacy of single-case and case series treatment designs and their proposals for modifications to present practices that will improve the validity and robustness of these methods. Although the review of currently used approaches is not exhaustive, it addresses some of the most troublesome roadblocks faced by aphasiologists who conduct single-case and case series treatment research.

References

- Beeson PM, Robey RR. Evaluating single-subject treatment research: Lessons learned from the aphasia literature. *Neuropsychology Review*. 2006; 16:161–169. [PubMed: 17151940]
- Dell GS, Schwartz MF, Martin N, Saffran EM, Gagnon DA. Lexical access in aphasic and nonaphasic speakers. *Psychological Review*. 1997; 104:801–838. [PubMed: 9337631]
- Elman R. Evidence-based practice: What evidence is missing? *Aphasiology*. 2006; 20:103–109.
- Fisher WW, Kelley ME, Lomas JE. Visual aids and structured criteria for improving visual inspection and interpretation of single-case designs. *Journal of Applied Behavior Analysis*. 2003; 36:387–406. [PubMed: 14596583]
- Franklin, S. Ranklin, RD. Allison, DB., Gorman, BS., editors. *Design and analysis of single-case research*. Mahwah, NJ: Lawrence Erlbaum Associates; 1997.
- Howard D. Forum: Evaluating intervention beyond randomised controlled trials: The case for effective case studies of the effects of treatment in aphasia. *International Journal of Language & Communication Disorders*. 1986; 21:89–102.
- Howard D, Best W, Nickels L. Optimising the design of intervention studies: Critiques and ways forward. *Aphasiology*. 2014 Advance online publication.
- Howard, D., Franklin, S. *Missing the meaning?*. Cambridge, MA: MIT Press; 1988.
- Levelt WJM, Roelofs A, Meyer AS. Multiple perspectives on word production. *Behavioral and Brain Sciences*. 1999; 22:61–69.
- McReynolds, LV., Kearns, KP. *Single subject experimental designs in communicative disorders*. Baltimore, MD: University Park Press; 1983.
- Patterson, KE., Shewell, C. Speak and spell: Dissociations and word-class effects. In: Coltheart, M. Job, R., Sartori, G., editors. *The cognitive neuropsychology of language*. Hove: Lawrence Erlbaum Associates; 1987.
- Rapp B, Goldrick M. Discreteness and interactivity in spoken word production. *Psychological Review*. 2000; 107:460–499. [PubMed: 10941277]
- Robey RR, Schultz MC, Crawford AB, Sinner CA. Review: Single-subject clinical-outcome research: Designs, data, effect sizes, and analyses. *Aphasiology*. 1999; 13:445–473.
- Rothi LJ, Barrett AJ. Introduction—The changing view of neurorehabilitation: A new era of optimism. *Journal of the International Neuropsychological Society*. 2006; 12:812–815. [PubMed: 17124756]
- Shewhart, WA. *Economic control of quality of manufactured products*. New York, NY: D. Van Nostrand; 1931.
- Swoboda, CM., Kratochwill, TR., Levin, JR. *Conservative dual-criterion method for single-case research: A guide for visual analysis of AB, ABAB, and multiple-baseline designs*. Madison: Center for Education Research, University of Wisconsin; 2010. (WCER Working Paper No. 2010-13). Retrieved from <http://www.wcer.wisc.edu/publications/workingPapers/papers.php>
- Thompson CK. Single subject controlled experiments in aphasia: The science and the state of the science. *Journal of Communication Disorders*. 2006; 39:266–291. [PubMed: 16635494]
- Whyte J, Gordon W, Rothi LG. A phased developmental approach to neurorehabilitation research: The science of knowledge building. *Archives of Medical Rehabilitation*. 2009; 90:S3–S10.