Supplemental Material S1. Modified systematic review protocol.

Administrative information

Title: Effect Sizes in Single Case Aphasia Studies: A Comparative, Autocorrelation-Oriented Analysis

Identification: The protocol presented here describes a selective meta-analysis conducted to generate data for assaying the defensibility of four commonly used effect size measures.

Authors:

Contact: Brent Archer, Department of Communication Sciences and Disorders, Bowling Green State University, OH

Jamie H. Azios, Department of Speech and Hearing Sciences, Lamar University, Beaumont, TX

Nicole Müller, Department of Speech and Hearing Sciences, University College Cork, Ireland

Lauren Macatangay, Department of Communication Sciences and Disorders, Bowling Green State University, OH

Correspondence concerning this article should be addressed to Brent Archer: barcher@bgsu.edu

Contributions: BA is the guarantor. BA, JA, and NM drafted the manuscript. All authors contributed to the development of the selection criteria, data extraction criteria and the search strategy. BA and LM searched the data bases, applied the search criteria and selected articles for further analysis. Data extraction was carried out by LM. BA provided statistical expertise. All authors read, provided feedback and approved the final manuscript.

Support

Sources: None

Sponsor: None

Introduction

Rationale: In single-case treatment studies, researchers may compare client performance during baseline, nontreatment phase(s) to client performance during intervention phases. Autocorrelation in the data series gathered during such studies increases the likelihood that analysts will detect or fail to detect meaningful differences between baseline and treatment phase data. We extracted time series data from previous studies of interventions designed to treat word retrieval deficits in people with aphasia (PWA) and examined the impact that autocorrelation has on four effect size calculation methods when these methods are applied to published data. We selected two older statistics, namely Busk and Serlin's *d* and Young's *C*, and two newer procedures designed to minimize the impact of autocorrelation, namely nonoverlap of all pairs (NAP) and Tau-*U*.

Objective: The aim of this selective meta-analysis was to furnish time series data generated by PWA in treatment studies for word retrieval deficits. These time series data were then subjected to further analysis to determine the extent to which the four effect size calculation methods outlined above are influenced by autocorrelation.

Methods

Eligibility criteria:

1. Only articles published in English were selected.
2. Articles had to describe an investigation of an intervention approach designed to improve word retrieval abilities in people with anomia associated with a neurogenic communication disorder.
3. Studies which provided strong, medium, weak, or no evidence for a given intervention approach were all included (i.e., studies which found that a given intervention approach was ineffective were also included).
4. We limited our search to articles published between 1985 and 2016.
5. Studies had to feature at least one time series or an array of data points which track a given behavior in a single research participant over time.
6. Studies had to be divided into at least two phases, namely a baseline phase in which participant naming abilities were assessed via formalized testing or some other numerically-based measure and a treatment phase during which the intervention was delivered. Studies that featured additional phases (e.g., follow-up) were also included, even though post-B phase data were not used for analysis.
7. The time series data points had to be retrievable from material published in the study.
8. Studies which featured time series of any length were included.

Information sources: Linguistics and Language Behavior Abstracts, Biosis Previews, CINAHL Plus, Health Source (Nursing/Academic Edition), PsychINFO (1887-current), PubMed, speechBITE, JSTOR, OhioLink Electronic Journal Center, Project MUSE, and Web of Science.

Search strategy: We classified search terms and phrases into two groups. Terms describing disorders were "adult communication deficit", "adult communication disorder", "adult communication impairment", "adult neurogenic communication deficit", "adult neurogenic communication disability", "adult neurogenic communication disorder", "adult neurogenic communication impairment", "anomia", "aphasia", "naming", "naming deficit", "naming difficulty", "naming disorder", "naming impairment", "word finding deficit", "word finding difficulty", "word finding disorder", "word finding impairment", "word retrieval deficit", "word retrieval difficulty", "word retrieval disorder", and "word retrieval impairment". The terms and phrases describing intervention aspects were "AB investigation", "AB study", "AB trial", "ABAB investigation", "ABAB study", "ABAB trial", "baseline", "intervention", "multiple baseline investigation", "multiple baseline study", "multiple baseline trial", "single case investigation", "single case quasi-experimental investigation", "single case quasi-experimental study", "single case quasi-experimental trial", "single case study", "single case trial", "single case", "small *n* investigation", "small *n* study", "small *n* trial", and "treatment".

Each of the terms and phrases from the disorder description group were paired with each of the terms and phrases from the intervention description group, which yielded 374 combinations.

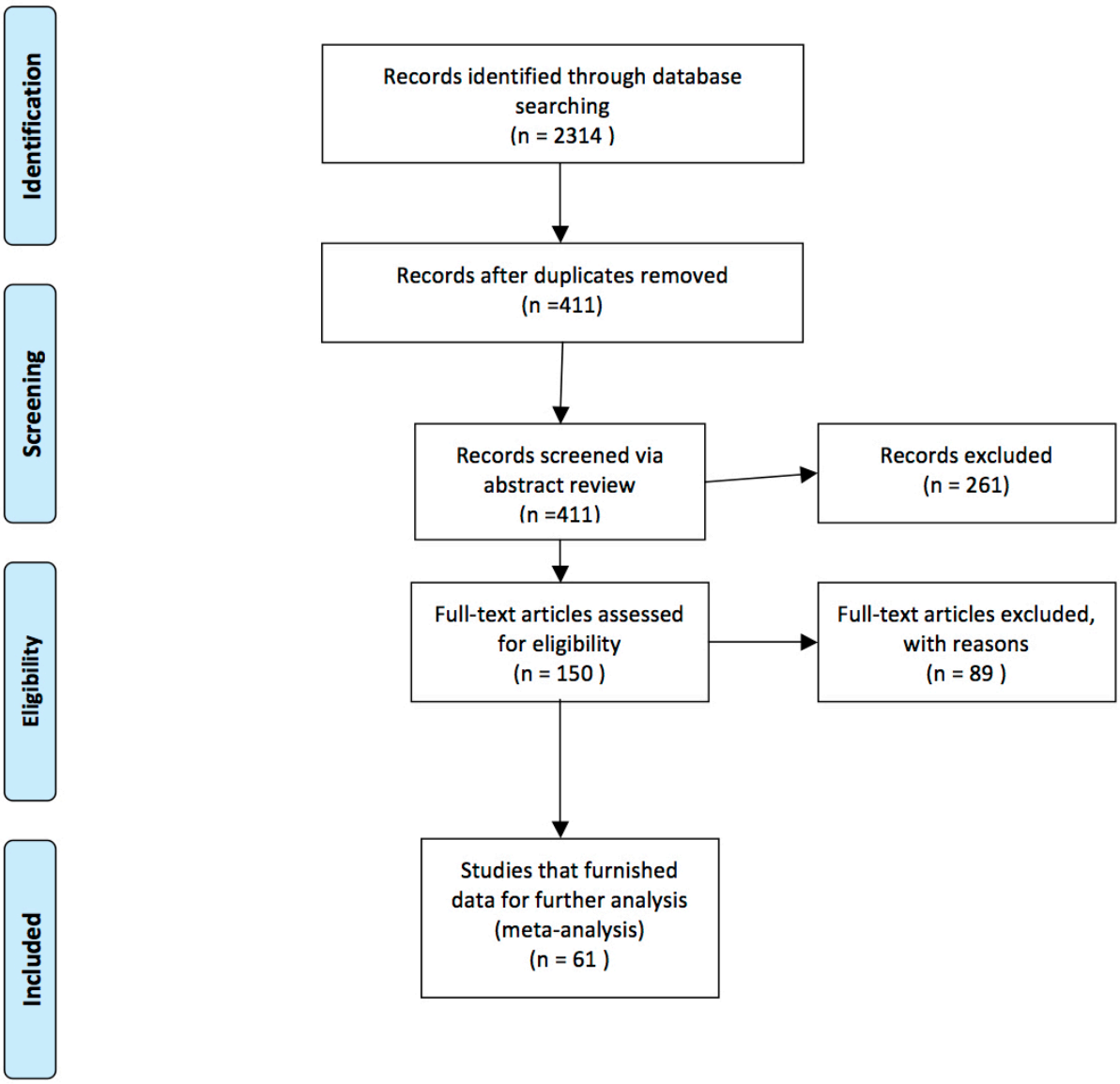
Study Records

Data management: Copies of article abstracts and full texts were stored on an online file hosting service (Google Docs) to which BA and LM had access. Time series data were entered into a simple database that consisted of a number of linked Microsoft Excel spreadsheets arranged into workbooks. These spreadsheets were also stored on Google Docs.

Selection process: Initial database searching was carried out by the BA and LM authors between August of 2016 and January of 2018. BA had more than a decade's clinical and academic experience within the field of aphasiology and had obtained a doctoral level qualification in communication disorders. LM was enrolled in a graduate (master's) speech pathology program at the time of the study. Our initial database search yielded 2,314 records. After duplicates were removed, the number of records was reduced to 411.

Abstracts for all of the 411 articles that met search criteria 1–4 were retrieved and reviewed by BA and LM. Based on the initial abstract review, BA and LM selected and retrieved the full texts of the articles that they thought would be most likely to meet criteria 5–8 (150 articles). Criteria 5–8 were then applied to select the articles that would yield further data for analysis (61 articles). Figure 1 provides a graphical illustration of the selection process.

Figure 1.



In cases where the BA and LM disagreed about whether or not a study should be included, the authors conferred and discussed whether or not the study in question fit the criteria described. Decisions about inclusion or exclusion were based on the consensus view reached in these discussions.

Data collection process: LM reviewed the articles and extracted baseline and treatment phase data points which were then used to create time series.

In order to ensure that data points were accurately extracted and entered into the time series database, a research assistant reviewed the articles and extracted baseline and treatment phase data. All of the 3,636 data points extracted by the fourth author and the research assistant were compared. The fourth author and the research assistant were in agreement for over 99.5% of the data points. Only 18 data points (less than 0.5%) were recorded differently by the two team members involved in the process of extraction. In each of these instances, the relevant studies were reviewed for a third time and the correct data points were included in the time series database.

Data items: For the purposes of this study, if the authors provided data concerning a behavior as displayed by a single participant over a defined period of time, it was included in the time series database. Since most of the studies featured numerous participants whose behavior was tracked under numerous conditions, a single study would often provide more than one time series. For example, Law, Wong, Sung, and Hon (2006) tracked two participants' ability to name items on two word lists (high and low familiarity). This study provided four time series (two participants X two word lists) that we entered into the time series database.

Outcomes and prioritization: Several studies included data series associated with a list of items that were directly treated in the intervention and data series associated with phonemically and/or semantically related items. In many instances, the related items were not directly treated but participants' naming performance for those lists were included if the authors were oriented to questions of generalization. For our analysis, we included both treatment and generalization data series.

While studies that aim to determine whether or not the effects of treatment persist beyond the conclusion of treatment (for example, Coehlo, McHugh, & Boyle, 2000) furnish findings that are highly clinically relevant to models of service delivery, at present a relatively small number of published articles provide this kind of follow-up data. The statistical analysis we carried out would not have been feasible with such a limited data set. Thus we chose not to include data from beyond the point at which the first treatment phase ended. Even if a given study included data from beyond the first treatment phase (e.g., follow-up data, data from later phases in ABABAB designs), this was not entered into the database.

Risk of bias in individual studies: As was the case Parker and Brossart (2003), Parker and Vannest (2009), Parker, Vannest, Davis, and Sauber (2011), Ross (2012), and Ross and Begeny (2014), our objective in conducting this selective meta-analysis was to furnish ourselves with data that could be used in a statistical analyses designed to examine how robust four effect size calculation methods are against the effects of autocorrelation. Accordingly, we were not concerned that the bias evident in individual studies might influence our final assessment of a treatment approach, since our goal was not to provide a systematic review of the efficacy of a given intervention or interventions.

Data

Synthesis: As referenced above, the objective of this selective meta-analysis was not to provide an overall assessment of the efficacy of a given intervention approach. It was thus not necessary for us to engage in a synthesis process.

Meta-biases: See above.

Confidence in cumulative evidence: See above.

Conversion formulae:

Converting *d* to *r* (Borenstein, Hedges, Higgins, & Rothstein, 2011)

$$r = \sqrt{\frac{d^2}{4 + d^2}}$$

Converting *C* to *r* (Tryon, 1982; Wolf, 1986)

$$SEM(C) = \sqrt{\frac{N + 2}{(N - 1)(N + 1)}}$$

$$Z = \frac{C}{SEM(C)}$$

$$r = \sqrt{\frac{Z^2}{Z^2 + N}}$$

Converting AUC to *d* (Ruscio, 2008)

$$d = 3.46(\sqrt{2 - 2(\mathrm{AUC})} + 1)$$

References

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. Chichester, United Kingdom: Wiley

Coelho, C. A., McHugh, R. E., & Boyle, M. (2000). Semantic feature analysis as a treatment for aphasic dysnomia: A replication. *Aphasiology*, *14*(2), 133–142.

Law, S.-P., Wong, W., Sung, F., & Hon, J. (2006). A study of semantic treatment of three Chinese anomic patients. *Neuropsychological Rehabilitation, 16*(6), 601–629.

Parker, R. I., & Brossart, D. F. (2003). Evaluating single-case research data: A comparison of seven statistical methods. *Behavior Therapy, 34*(2), 189–211.

Parker, R. I., & Vannest, K. (2009). An improved effect size for single-case research: Nonoverlap of all pairs. *Behavior Therapy, 40*(4), 357–367.

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining nonoverlap and trend for single-case research: Tau-U. *Behavior Therapy, 42*(2), 284–299.

Ross, S. (2012). *Measuring response to intervention: Comparing three effect size calculation techniques for single-case design analysis* (Unpublished doctoral dissertation). North Carolina State University, Raleigh, NC.

Ross, S., & Begeny, J. C. (2014). Single-case effect size calculation: Comparing regression and non-parametric approaches across previously published reading intervention data sets. *Journal of School Psychology*, *52*(4), 419–431.

Ruscio, J. (2008). A probability-based measure of effect size: Robustness to base rates and other factors. *Psychological Methods*, *13*(1), 19–30.

Tryon, W. W. (1982). A simplified time-series analysis for evaluating treatment interventions. *Journal of Applied Behavior Analysis, 15*(3), 423–429.

Wolf, F. M. (1986). *Meta-analysis: Quantitative methods for research synthesis*. New York, NY: Sage.