Quality of Systematic Reviews of Treatment Studies in Neurogenic Communication Disorders

Evidence based clinical practice involves the blending of clinical expertise, patient values, and current best research evidence (Sackett et al.,1996). Clinicians report that the greatest challenge in implementing evidence based practice is finding time to remain abreast of the current best clinical evidence (Mullen, 2005). Several efforts have been initiated to accumulate clinical evidence in the form of systematic reviews of the treatment literature, as in the ANCDS Practice Guidelines project, ASHA National Center for Evidenced-based Practice, and the Cochrane Database.

Because of the importance of systematic reviews to evidence based practice, it is essential that reviews be conducted with rigorous methodologies to avoid bias in conclusions (Schlosser et al, 2007). Several methods have been proposed to allow for the evaluation of the quality of systematic reviews (Dolloghan, 2007; Schlosser et al., 2007); as of yet there is no established standard for appraisal of reviews. In general, a systematic review should include an established protocol to address a set of clinical questions. All sources should be identified and selection criteria established. A procedure for summarizing the data and evaluating the quality of the studies also is necessary. Reviews that include meta-analysis should include a variety of statistic methods. Among the strongest appraisal criteria are those described by Auperin et al. (1997). The purpose of this paper is to report on our evaluation of systematic reviews that have been conducted for treatment studies in neurogenic communication disorders.

Methodology: We identified 15 systematic reviews of behavioral treatment research in neurogenic communication disorders (aphasia n=6, apraxia of speech n=3, dysarthria n=6) through searches of databases (PubMed, ANCDS, Cochrane) and hand searches of several journals. Two reviewers independently rated each systematic review on a set of 27 criteria described by Auperin et al. (1997) and rated on a scale of 0 (not included), 1 (addressed partially), and 2 (addressed adequately), leading to a maximum score of 54. Criteria examined the identification of protocol and selection of trials for each review (6 items), description of the clinical trials (4 items), evaluation of study quality (3 items), description of data collection procedures (3 items), statistical analysis (8 items), and application of results (3 items). The raters trained on the use of the coding system with two of the 15 articles (Robey, 1998; Palmer & Enderby, 2007). A third examiner was consulted when a discrepancy in scoring was identified. Upon resolution of discrepancies for these two articles, all other articles were coded independently. Coding agreement was at 97%.

Results: Quality scores for the 15 articles are shown in Table 1. Among the 6 aphasia articles (3 systematic reviews and 3 meta-analyses), quality scores ranged from 8 (Holland et al., 1996) to 42 (Greener et al., 1999), with a mean score of 25.5/54. Of the 6 dysarthria articles (all 6 systematic reviews), quality scores ranged from 10 (Palmer & Enderby (2007) to 33 (Deane et al., 2001), with a mean score of 22.17/54. Among the 3 apraxia of speech articles (2 systematic reviews and 1 meta-analysis), scores ranged from 16 (Wambaugh & Doyle, 1994) to 47 West et al. (2008), with a mean score of 27.6/54.

Several of the rating criteria were weighted in favor of statistical analyses that are typically seen in meta-analysis. Therefore we evaluated the studies on a sub-score for those criteria relevant to

systematic reviews, leading to a maximum score of 38. The mean scores were: 20.50 aphasia; 22.17 dysarthria, and 23.00 apraxia of speech. That is, apraxia of speech reviews tended to be of higher quality whether using the full 54 point score or the modified 38 point score, largely skewed by an excellent review by West et al. (2008). Aphasia systematic reviews tended to have lower scores overall.

Table 2 displays the number of studies that 'adequately addressed' each of the 27 different criteria evaluated. Only 8 criteria were met by a majority of the 15 reviews, largely focusing on the methods for identifying the studies and descriptions of the studies included in the reviews. Several of the criteria important for avoiding bias in the synthesis and conclusions of the reviews were lacking. Other weaknesses across reviews centered on lack of statistical analyses.

Discussion: Some strong reviews have been completed in treatment studies for neurogenic communication disorders, particularly those conducted under the auspices of the Cochrane Collaboration. Yet there is considerable room for improvement across reviews. Some lower scores were noted for studies conducted in the more distant past (Holland et al., 1996; Wambaugh & Doyle, 1994) when methods for the conduct of systematic reviews were not as well established. Clearly, methods employed in systematic reviews and meta-analyses have evolved over the past two decades leading to an overall improvement in the quality of reviews in recent years. Nevertheless, there is still room for improvement. This project demonstrated patterns of weakness across reviews, including the observation that meta-analytic methods are too rarely implemented. In both systematic reviews and meta-analyses, researchers need to improve their reporting of protocols, extraction procedures to avoid selection bias, and consideration of the economic impact of the treatment research findings.

The importance of current best evidence in evidence-based practice continues to receive primary consideration in medical speech pathology circles. Therefore, efforts will continue in the completion of systematic reviews and meta-analyses for neurogenic communication disorders. Our study is meant to challenge those completing such work to use rigorous review methods such that clinicians have access to summaries of the best, non-biased clinical research evidence.

Table 1: Quality scores in review articles.

| Article | Score Max 54 | Modified Score Max 38 |
|---|---|---|
| *Aphasia* | | |
| Holland et al., 1996 | 8 | 8 |
| Robey, 1998 | 30 | 22 |
| Greener et al., 1999 | 42 | 28 |
| Bhogal et al., 2003 | 28 | 20 |
| Turner & Whitworth, 2006 | 18 | 18 |
| Cherney et al., 2008 | 27 | 27 |
| *Dysarthria* | | |
| Deane et al., 2001 (PKN) | 30 | 30 |
| Deane et al., 2001 (placebo) | 33 | 33 |
| Yorkston et al., 2001 | 19 | 19 |
| Yorkston et al., 2003 | 20 | 20 |
| Palmer & Enderby, 2007 | 10 | 10 |
| Yorkston et al., 2007 | 21 | 21 |
| *Apraxia of Speech* | | |
| Wambaugh & Doyle, 1994 | 16 | 16 |
| Wambaugh et al., 2006 | 20 | 20 |
| West et al., 2008 | 47 | 33 |

Table 2: Number of studies achieving an ‘adequately addressed’ score for each methodologic criterion.

| Criterion | Aphasia (n=6) | Dysarthria (n=6) | Apraxia of Speech (n=3) | Total (n=15) |
|---|---|---|---|---|
| Protocol | 0 | 1 | 0 | 1 |
| Literature search | 4 | 5 | 3 | 12 |
| List of trials analyzed | 6 | 6 | 3 | 15 |
| Log of rejected trials | 2 | 1 | 0 | 3 |
| Selection method | 1 | 0 | 0 | 1 |
| Control of publication bias | 3 | 2 | 1 | 6 |
| Description of patients | 4 | 3 | 2 | 9 |
| Description of treatments | 4 | 6 | 3 | 13 |
| Description of diagnoses | 4 | 6 | 3 | 13 |
| Clinical combinabilitiy criteria | 3 | 0 | 1 | 4 |
| Only randomized trials pooled | 1 | 2 | 1 | 4 |
| Trial quality assessment | 5 | 4 | 3 | 12 |
| Intention-to-treat analysis | 1 | 1 | 1 | 3 |
| Data extraction method | 2 | 0 | 1 | 3 |
| Inter-observer agreement | 3 | 2 | 2 | 7 |
| Contact with trial investigators | 1 | 2 | 1 | 4 |
| Statistical methods | 3 | 0 | 1 | 4 |
| Statistical errors | 0 | 0 | 0 | 0 |
| Confidence intervals | 2 | 0 | 1 | 3 |
| Test of homogeneity | 1 | 0 | 1 | 2 |
| End point quality | 3 | 0 | 1 | 4 |
| Sensitivity analysis | 1 | 0 | 1 | 2 |
| Subgroup analyses | 2 | 0 | 1 | 3 |
| Indirect analyses | 3 | 0 | 1 | 4 |
| Clinical impact | 5 | 5 | 2 | 12 |
| Economic impact | 0 | 2 | 0 | 2 |
| Specification of source of support | 5 | 5 | 3 | 13 |