This is the *Accepted Manuscript* of an article published by the American Speech-Language-Hearing Association (ASHA) in the *Journal of Speech, Language, and Hearing Research, 61*, 2757-2771 © 2018. The manuscript is reprinted here with permission from ASHA and is further available online https://pubs.asha.org/doi/10.1044/2018_JSLHR-S-17-0366

Minimally Detectable Change and Minimal Clinically Important Difference of a Decline in Sentence Intelligibility and Speaking Rate for Individuals with ALS

Kaila L. Stipancic, M.A., CCC-SLP[1], kstipancic@mghihp.edu, (617) 724-6347

Yana Yunusova, Ph.D., CCC-SLP[2], yana.yunusova@utoronto.ca, (416) 946-5456

James D. Berry, M.D., M.P.H.[3], jdberry@mgh.harvard.edu, (617) 724-3914

Jordan R. Green, Ph.D., CCC-SLP[1], jgreen2@mghihp.edu, (617) 945-3624

[1]*Department of Communication Sciences and Disorders, MGH Institute of Health Professions, Boston, MA, United States*

[2]*Department of Speech-Language Pathology, University of Toronto, Toronto, Canada*

[3]*MDA ALS Clinic, Massachusetts General Hospital, Boston, MA, United States*

Corresponding author: Jordan R. Green
MGH Institute of Health Professions
Building 79/96, 2nd floor
13th Street
Boston, MA 02129
(617) 945-3624
jgreen2@mghihp.edu

Conflict of Interest:
There are no relevant conflicts of interest to disclose.

Funding:
R01DC0135470 (PI: Jordan R. Green)
R01DC009890 (PIs: Jordan R. Green, Yana Yunusova)
K24DC016312 (PI: Jordan R. Green)

Minimally Detectable Change and Minimal Clinically Important Difference of a Decline in Sentence Intelligibility and Speaking Rate for Individuals with ALS

Abstract

Purpose: The purpose of this study was to determine the Minimally Detectable Change (MDC) and Minimal Clinically Important Difference (MCID) of a decline in speech sentence intelligibility and speaking rate for individuals with amyotrophic lateral sclerosis (ALS). We also examined how the MDC and MCID vary across severities of dysarthria.

Method: One-hundred and forty-seven patients with ALS and 49 healthy control subjects were selected from a larger, longitudinal study of bulbar decline in ALS, resulting in a total of 650 observations. Intelligibility and speaking rate in words per minute (WPM) were calculated using the Sentence Intelligibility Test (Yorkston, Beukelman, & Hakel, 2007) and the ALS Functional Rating Scale – Revised (Cedarbaum et al., 1999) was administered to capture patient perception of motor impairment. The MDC at the 95% confidence level was estimated using the following formula: $MDC_{95} = 1.96 \times \sqrt{2} \times SEM$. For estimation of the MCID, receiver operating characteristics curves were generated and area under the curve and optimal thresholds to maximize sensitivity and specificity were calculated.

Results: The MDC for sentence intelligibility was 12.07% and the MCID was 1.43%. The MDC for speaking rate was 36.57 WPM and the MCID was 8.80 WPM. Both MDC and MCID estimates varied with severity of dysarthria.

Conclusions: The findings suggest that declines greater than 12% sentence intelligibility and 37 WPM are required to be outside measurement error, and that these estimates vary widely across dysarthria severities. The MCD and MCID metrics used in this study to detect real and clinically relevant change should be estimated for other measures of speech outcomes in intervention research.

Introduction

Speech intelligibility, or how understandable a speaker is to a listener, is a common clinical metric of communication effectiveness in persons with motor speech impairments (Cannito, Suiter, Beverly, Chorna, Wolf, & Pfeiffer, 2012; Kent, Weismer, Kent, & Rosenbek, 1989; Yorkston & Beukelman, 1981). Improving intelligibility is often the primary goal of speech-language therapy for these individuals (Miller, 2013), with the ultimate objective of positively impacting quality of life. Another widely used measure of speech motor involvement is speaking rate, as measured in words per minute (WPM). For progressive diseases, such as amyotrophic lateral sclerosis (ALS), speaking rate is a preferred marker of early speech motor involvement because it declines earlier in the disease than intelligibility (Ball, Beukelman, & Pattee, 2002; Green et al., 2013). Although clinicians and researchers can track changes in speech intelligibility and speaking rate over time, the field is currently lacking critical information regarding the magnitude of change in these outcome measures that is necessary to be considered clinically meaningful. This information is essential for determining both the positive effects of speech therapy and the detrimental effects of neurologic disease progression on speech. Herein, we provide a review of applications of detectable and clinically meaningful change, followed by a study exploring the application of these concepts to speech outcomes in individuals with ALS.

Standard Framework for Estimating Detectable and Clinically Meaningful Change

The minimally detectable change (MDC) and the minimal clinically important difference (MCID) have been used extensively in other fields, such as such as physical therapy (Beninato, Fernandes, & Plummer, 2014; Riddle & Stratford, 2013), occupational therapy (Coster, 2013; Wu, Chuang, Lin, Lee, & Hong, 2011), and healthcare (Jaeschke, Singer, & Guyatt, 1989), to

estimate real and important change, respectively (Streiner & Norman, 2008). Although the MDC and MCID are related, they are clearly distinct terms that define separate attributes of a particular measure.

Minimally detectable change (MDC). The MDC has been defined as the smallest amount of change that is greater than measurement error (Beckerman, Roebroeck, Lankhorst, Becher, Bezemer, & Verbeek, 2001; de Vet, Terwee, Ostelo, Beckerman, Knol, & Bouter, 2006; Haley & Fragala-Pinkham, 2006). In other words, if change occurs that is outside the normal variation of a measurement on repeated trials, we can be confident that this change is not simply due to random variability (Riddle & Stratford, 2013). The MDC is often calculated by obtaining a reliability statistic of the measurement in question on patients who have not clinically changed (Fulk & Echternach, 2008; Haley & Fragala-Pinkham, 2006; Kovacs et al., 2008; Lehman & Velozo, 2010; Mallinson, Pape, & Guernon, 2016; Riddle & Stratford, 2013). The MDC threshold is set using a confidence level. A 95% confidence level indicates that there is only a 5% chance that a change above this threshold could be due to chance variability in a truly unchanged patient (Stratford & Riddle, 2012). While a change greater than the MDC indicates that the change is unlikely to be due to chance variability, it does not indicate whether this degree of change is clinically meaningful. (Beninato & Portney, 2011). For this reason, the MCID is a necessary supplement to the MDC.

Minimal clinically important difference (MCID). The MCID has been defined as the smallest amount of change in a domain (e.g., balance, gait, quality of life, pain, etc.) that is considered relevant or important to patients, clinicians, or significant others (de Vet et al., 2006; Hays & Woolley, 2000; Jaeschke et al., 1989). Although, the concept of the MCID was initially developed to appraise the significance of improvements in function, it has recently been applied

to declines in function, such that the MCID can indicate the smallest amount of change which patients would perceive as detrimental. Identifying the MCID of a novel metric requires an external standard of meaningfulness (Jaeschke et al., 1989). Ideally, the external anchor is a gold-standard assessment (Coster, 2013; Haley & Fragala-Pinkham, 2006). The threshold for clinical meaningfulness on the gold-standard scale is based on clinical acumen (i.e., the clinician or researcher often decides how many points or levels are required for 'important change' to have occurred) and the corresponding score on the new metric is considered the MCID. The MCID is an estimate of important change, and must be larger than the MDC to be useful (Beninato et al., 2014).

Example of the MDC and MCID in the occupational therapy literature. A 2011 study provides an illustration of how these constructs have been used in the occupational therapy literature (Wu et al., 2011). The authors were interested in estimating the MDC and MCID of the Nottingham Extended Activities of Daily Living (NEADL; Nouri & Lincoln, 1987) scale. The NEADL scale is a 22-point scale that measures patients' independence in four areas of daily life: mobility, kitchen, domestic, and leisure activities on a scale from 0 (unable to complete) to 3 (able to complete). Using methods similar to ours (outlined below), the authors calculated the MDC (4.9 points) and the MCID (6.1 points) for the NEADL scale. The authors concluded that both the MDC and MCID are important for clinicians to use when determining if the changes in their patients' independence in daily activities are real and/or relevant, and to ultimately aid in therapeutic management decisions.

Applying the MDC and MCID to Speech Outcomes

Although the MDC and MCID have been used widely in other fields and are deemed important for use by clinicians and researchers (Beninato et al., 2014; de Vet et al., 2006;

Jaeschke et al., 1989; Kovacs et al., 2008; Revicki, Hays, Cella, & Sloan 2008), these concepts have not, to our knowledge, been applied to speech outcomes. Application of the MDC and MCID to speech outcomes such as intelligibility and speaking rate is critically important not only for setting appropriate expectations regarding treatment outcomes, but also for knowing when these treatments are making real and clinically important relevant changes for patients. A sampling of previously reported treatment effects on percent speech intelligibility within the dysarthria literature reveals a wide range of gains from 2 to 33 percentage points (see Table 1).

[Table 1]

The table demonstrates the variety of populations, therapies, and metrics of intelligibility in the literature. Although a number of studies have reported positive effects of speech treatment on intelligibility in people with dysarthria, provision of the associated MDCs and MCIDs could increase confidence that these therapies yield real and meaningful changes. Incorporating these concepts into speech outcome research will help better define disease progression in individuals whose speech intelligibility and speaking rate are declining and also provide clarity about impending needs for augmentative communication strategies. In addition, standardizing this lexicon in speech outcomes research will facilitate clear communication between researchers, clinicians, and patients who would all benefit from an index with which to judge the clinical importance of measured changes.

Challenges to estimating the MDC and MCID of sentence intelligibility and speaking rate. Despite the potential benefits of establishing MDC and MCID for speech outcomes, there are several significant challenges to applying these concepts to speech intelligibility and speaking rate. First, there are no gold standard measures of clinically meaningful change in speech intelligibility or speaking rate. There are limited psychometrically

validated metrics of patient, caregiver, and clinician perceptions of clinically important change, and those that exist ask about concepts such as speaker/listener effort, speech naturalness, and comprehensibility, and are not directly related to intelligibility (see Baylor, Yorkston, Eadie, Miller, & Amtmann, 2009; Donovan, Kendall, Young, & Rosenbek, 2008; Hartelius, Elmberg, Holm, Lövberg, & Nikolaidis, 2008; Walshe, Peach, & Miller, 2009). Furthermore, the MCID may differ by (1) perspective (i.e., patients, communication partners, or clinicians), (2) direction of change (i.e., whether patients are getting better or worse) (Hays & Woolley, 2000), and (3) baseline speech impairment (Beninato, Gill-Body, Salles, Stark, Black-Schaffer, & Stein, 2006; Wang, Hart, Stratford, & Mioduski, 2011). More severely impaired patients may require a greater change to reach clinical relevance (Kovacs et al., 2008; Riddle & Stratford, 2013). Although, typically, the MDC and MCID have been reported as a single score, an exploration of how MDC and MCID estimates vary across dysarthria severities was of interest for the current study.

Calculating the MDC. Recall that in order to estimate the MDC, reliability and the error rate of the measure in question are needed. Although the MDC has not been explicitly estimated for speech intelligibility and speaking rate, these two necessary components for calculating the MDC (reliability and error rate) have been previously reported for speech intelligibility.

***Reliability*.** In the current study, we investigated sentence intelligibility as measured by orthographic transcription, or a word-for-word writing of what a listener thinks a speaker said, yielding percentage of words correctly transcribed. Additionally, since the MDC is typically calculated using test-retest reliability, presumably from a single judge using the same metric on two occasions, we explored intralistener reliability, a single judge using the same metric on one occasion, as a corollary. Several authors have previously reported the intralistener reliability on

the orthographic transcription of dysarthric speech with correlation coefficients ranging from .58 to 1.00 (mean = .80, SD = .13) (Tjaden, Kain, & Lam, 2014) and from .32 to .88 (mean = .66, SD = .13) (Stipancic, Tjaden, & Wilding, 2016). However, in these two studies, listeners were presented stimuli that were mixed with multitalker babble. Therefore, these reliability statistics should be cautiously compared to protocols in which listeners were presented with stimuli free of background noise, which have reported intralistener reliability with correlation coefficients ranging from .83 to .99 (Bunton, Kent, Kent, & Duffy, 2001; Hustad, 2006; Keintz, Bunton, & Hoit, 2007; Yorkston & Beukelman, 1978; 1981). Stratford (2004) argued that a correlation coefficient alone is not suitable for assessing the confidence (or "realness") of a measured value. Because reliability statistics are not in the same units as the measurement in question, their utility for meaningfully evaluating what a patient's score on the measure indicates is limited (Stratford, 2004). For this type of evaluation to be possible, the reliability estimates must be accompanied by the standard error of measurement to allow for calculation of the MDC. Variable reliability of intelligibility measurement across speech severities (Hustad, Oakes, & Allison, 2015) further supports an exploration of MDC and MCID change across dysarthria severities.

Error rate. Keintz et al. (2007) reported reliability (r = .93) along with standard error of measurement (SEM = 1.44) of an orthographic transcription task. These statistics can be used to derive an MDC; however, the authors did not use these estimates to describe detectable or important change, which was not an objective of their study. In addition, an eight percent error rate of a visual analog scale (VAS) used for estimating the intelligibility of dysarthric speakers was previously found (Van Nuffelen, De Bodt, Vanderwegen, Van de Heyning, & Wuyts, 2010). The authors concluded that an "increase in intelligibility of >8% can be interpreted as clinically significant" (Van Nuffelen et al., 2010, p. 111). Characterizing this estimate as "clinically

significant" provides a somewhat incomplete picture because it is based solely on error rate and does not attempt to account for what is important to patients or listeners. This is a good example of how the concepts of the MDC and MCID can facilitate the usage of a universal language by disambiguating misused terminology in the speech outcome literature. Overall, the MDC and MCID have the potential to add to our current understanding of the reliability and error of intelligibility metrics.

A Case to Apply MDC and MCID of Speech Intelligibility and Speaking Rate: Amyotrophic Lateral Sclerosis

In the current study, we estimated the MDC and MCID based on data obtained from a cohort of individuals with amyotrophic lateral sclerosis (ALS). ALS is a quickly progressing neurodegenerative disease that results in deterioration of speech motor control across the speech subsystems, resulting in a dysarthria. In the clinical setting, speech decline is typically monitored using patient reported scales (e.g., the bulbar subscore of the ALS Functional Rating Scale – Revised (ALSFRS-R; Cedarbaum et al., 1999)). Some clinics also track speech intelligibility decline and speaking rate reduction, largely by clinician estimates of these perceptual constructs in conversation; however, MDC and MCID have not been defined for speech decline in ALS.

In this study, we applied the concepts of the MDC and MCID to ALS speech outcomes, to provide a framework for subsequent endeavors to estimate detectable and clinically relevant changes due to speech-language therapy and disease progression, as well as within other measures, contexts, and populations. To this end, the following aims were the focus of this study:

1. Determine the minimally detectable change (MDC) of a decline in sentence intelligibility and speaking rate on the Sentence Intelligibility Test (SIT; Yorkston, Beukelman, & Hakel, 2007) for individuals with ALS.

2. Determine the minimal clinically important difference (MCID) of a decline in sentence intelligibility and speaking rate on the SIT for individuals with ALS.
3. Examine how the MDC and MCID vary across dysarthria severity.

Methods

Participants

We included 147 people with ALS and 49 controls with SIT data available from a larger, longitudinal study of bulbar decline in ALS (Green et al., 2013). Control participants were required to speak English as their primary language, have had no history of speech, language, or hearing problems, or neurological disease, and have normal hearing and adequate vision and literacy skills to read stimuli. Participants with ALS had been diagnosed with ALS by a neurologist based on the El Escorial criteria (Brooks, Miller, Swash, & Munsat, 2000) and met the following inclusion criteria: 1) spoke English as their primary language, 2) had no prior history of neurological disorders, and 3) had normal hearing and adequate vision and literacy skills to read stimuli. The site of onset (i.e., bulbar vs. spinal) for participants with ALS varied, as did the severity of bulbar symptoms. The participants had a wide range of severity of bulbar symptoms with intelligibility scores ranging from 2.73% to 100% (mean = 93.35%, SD = 16.00) and speaking rates ranging from 37.03 WPM to 284.30 WPM (mean = 62.31 WPM, SD = 46.21) as measured by the Sentence Intelligibility Test (SIT; Yorkston, Beukelman, & Hakel, 2007). For comparison, previous research has reported intelligibility in healthy individuals from 96-100% and speaking rate from 160-200 WPM (Rong, Yunusova, Wang, & Green, 2015; Shellikeri et al., 2016). Participants with ALS were assessed longitudinally (average number of sessions per participant = 4.3, SD = 3.8, minimum number of sessions per participant = 1, maximum number of sessions per participant = 18) with an average of 111 days between

sessions (SD = 85.80, min = 14, max = 1057). Data from multiple sessions were included in this analysis, resulting in a total of 650 data collection sessions, which were analyzed as unrelated (see similar methods in Rong et al., 2015; 2016). From this point on, when we refer to 'N', we are referring to the number of sessions.

Procedures

All participants completed a standard research protocol that included a measure of speech intelligibility, a measure of patient perceived impairment, and instrumentation-based measures of speech that were designed to capture function of the speech subsystems (Rong et al., 2015; Yunusova, Green, Wang, Pattee, & Zinman, 2011). Of interest for calculations of both the MDCs and MCIDs were sentence intelligibility and speaking rate. A measure of patient perceived changes in function (ALSFRS-R speech subscore) that was used to anchor meaningfulness of change was of interest for the calculations of the MCIDs.

Sentence intelligibility and speaking rate. During each data collection session, the Sentence Intelligibility Test (SIT; Yorkston et al., 2007) was administered to calculate sentence intelligibility and speaking rate. One naïve adult listener (research assistant) who was unfamiliar with both the test materials and the speech/severity of the participants, orthographically transcribed the sentences produced by each speaker from audio recordings. Only one listener, rather than multiple listeners, was used to maximize the ecological validity of our MDC and MCID estimates; in a typical ALS clinical setting, only a single listener transcribes the speech of a patient with dysarthria. The listener was seated at a computer in a quiet lab environment and was able listen to each sentence twice before transcribing each sentence to the best of their ability. The listener heard all 11 sentences from each subject in order, as is standard for transcription of the SIT. Multiple listeners transcribed the speech samples over the time span of

the study (2009 to 2017). Because the sentence list was randomly generated for each participant from a large inventory, it was unlikely that the same sentence was used frequently enough for the listeners to become familiar with it. This listening protocol has been used in a number of previous studies (Rong, Yunusova, Wang, & Green, 2015; Yunusova et al., 2010; Yunusova, Green, Wang, Pattee, & Zinman, 2011). Percent intelligibility for each sentence (sentence intelligibility = number of words correctly identified in sentence/ number of words in sentence *100) and overall intelligibility (overall intelligibility = number of total words correctly identified across all sentences/ total words produced *100) were calculated. Speaking rate was calculated by dividing the number of words produced by the total utterance duration in minutes to yield a rate for each sentence in words per minute (WPM). For each speaker, an overall speaking rate was determined by averaging the speaking rates each all utterances.

External anchor of meaningfulness. The ALSFRS-R (Cedarbaum et al., 1999) was administered to the participants with ALS to measure patient perception of motor dysfunction. This highly reliable (Kaufmann et al., 2007) provider-guided scale is the current gold standard for measuring overall ALS severity and is widely used by speech-language pathologists and neurologists (Cedarbaum et al., 1999). The ALSFRS-R has been described as “the most widely accepted outcome measure of activity limitation for people with ALS” (Paganoni, Cudkowicz, & Berry, 2013, p. 608) and has been used as a primary outcome measure in many clinical trials (Gordon et al., 2007; Smith et al., 2017; Traynor et al., 2004). Additionally, the ALSFRS-R has been shown to be a useful predictor of ALS rate of progression and survival (Gordon et al., 2007; Kimura et al., 2006; Kollewe et al., 2008) and has been shown to be correlated with changes in strength and quality of life measures (Cedarbaum et al., 1999; Smith et al., 2017). The 12-item scale includes three items on bulbar function, one of which pertains to speech:

1. How is your speech?

    4 Normal speech process

    3 Detectable speech disturbance

    2 Intelligible with repeating

    1 Speech combined with non-vocal communication

    0 Loss of useful speech

Because this question is the primary indication of speech motor decline in many neurology clinics, this one question was used as an external standard with which to anchor the meaningfulness of change in sentence intelligibility and speaking rate for calculation of the MCID. Previous studies using the ALSFRS-R have similarly used single or multiple questions alone as outcome measures (Plowman et al., 2016; Smith et al., 2017). From this point on, when we refer to the ALSFRS-R subscore, we are referring only to the speech question of this scale. A one-point change on this question was utilized as an initial cut-off score based on previous research involving the full scale (see Castrillo-Viguera, Grasso, Simpson, Shefner, & Cudkowicz, 2010; Gordon et al., 2007). In addition, we also used a two-point cut off for determination of the MCID of intelligibility and speaking rate.

Dysarthria Severity Classification

To determine if MDC and MCID differed across severities, it was necessary to classify participants into severity groups. Although several studies have reported grouping participants by severity of intelligibility deficits (Blaney & Hewlett, 2007; Connaghan & Patel, 2017; dos Santos Barreto & Zazo Ortiz, 2016; Doyle et al., 1997; Hustad, 2006; 2007; 2008; Kent et al., 1989; Kim, Martin, Hasegawa-Johnson, & Perlman, 2010), these authors used widely varying categories and these categories were based on differing criteria (e.g., sentence intelligibility on

the SIT vs. scaled intelligibility of a conversation sample or reading passage) and different populations (e.g., cerebral palsy, PD, etc.). This motivated our use of two different stratification schemes for examining the effect of baseline intelligibility on the MDC: (1) a clinical weighting of dysarthria severity based on clinical acumen of SIT intelligibility scores and (2) an unbiased stratification of scores based on quantiles of the SIT intelligibility scores, which is presumably assumption free (see Figure 1). The clinical weighting of severity was accomplished by consulting a number of the papers previously cited and using our own clinical judgment to define severity stratifications.

[Figure 1]

Similarly, for the MDC of speaking rate, (1) a clinical weighting stratification was based on the divisions used by Shellikeri and colleagues (2016) and (2) an unbiased classification was based on quantiles of speaking rates as measured by the SIT to ensure that a different stratification of participants did not create significantly different results (see Figure 1). The MDC was then estimated for each of the strata across each of the classification schemes. The number of participants is slightly different across the MDC analyses, which can be seen in Figure 1. The number of subjects in each MDC analysis was restricted by the number of data collection sessions, as well as missing data for some of the variables. For example, first, we calculated the total MDC for speech intelligibility based on 639 observations (i.e., data collection sessions). Next, we calculated the MDC for speech intelligibility for each severity group. Because our groups were so heavily weighted to the normal end of intelligibility, it was necessary to subsample the full data set to ensure that the number of subjects in each severity group were similar. Cases within each group were selected using a random number generator.

Further, to examine the impact of dysarthria severity on the MCID, we stratified participants who had a one-point change on the ALSFRS-R subscore into groups based on baseline severity (i.e., initial ALSFRS-R subscore). We stratified the sample into three groups (see Figure 2): (1) patients with an ALSFRS-R subscore of four on the ALSFRS-R speech question at one session and a subscore of three at the next session (N = 63) who were compared to patients with a subscore of four at adjacent sessions (i.e., patients who were of the same severity, but were unchanged from one session to the next; N = 182), (2) patients with an ALSFRS-R subscore of three at one session and a subscore of two at the next session (N =18) who were compared to patients with a subscore of three at adjacent sessions (N = 80) and, (3) patients with an ALSFRS-R subscore of two at one session and a subscore of one at the next session (N = 3) who were compared to patients with a subscore of two at adjacent sessions (N = 14) (see Figure 2). In our sample, there were no patients with a subscore of one at one session and then a subscore of zero at an adjacent session. The MCIDs for intelligibility and speaking rate were estimated for each of these groups. The number of participants in each analysis can be seen in Figure 2.

[Figure 2]

Measures and Data Analysis

Minimally detectable change. Data from both the participants with ALS and control participants was used for calculation of the MDC. As mentioned previously, three parameters are required to estimate the MDC of a measure: 1) the difference between scores collected on the same or two different occasions, 2) reliability of the measure, and 3) the confidence level of interest. To derive the first two parameters, data is typically obtained from two separate data collection sessions that are minimally separated in time to ensure that patients are ‘unchanged’

from one session to the next. Because ALS is a quickly progressing disease, and two data collection sessions in close proximity to each other was not possible with our sample, we propose that the split-halves reliability of a measure will meet the first two requirements. Although test-retest reliability is the most commonly used statistic for calculating the MDC, internal consistency statistics, of which split-half reliability is an example, are also frequently used to estimate the MDC of a measure (de Vet et al., 2006; Nunnally & Bernstein, 1994). Furthermore, split-half reliability is typically higher than test-retest reliability. Therefore, use of split-half reliability provided an MDC that was estimated in a "best case scenario" or under the best circumstances. To calculate the split-half reliability of the SIT, the 11 sentences spoken by participants were divided into two halves. A total of seven random divisions of the SIT sentences were tested to ensure a stable estimate of split-half reliability. This dataset comparison was conducted to establish that the average intelligibility scores on each half of the SIT were not influenced by factors that vary systematically across sentences, such as sentence length. The mean difference and Pearson product correlations for the multiple split-half configurations are presented in Table 2.

[Table 2]

Correlation coefficients of the intelligibility scores on the two halves of these divisions ranged from 0.92 to 0.95. Additionally, we calculated the absolute mean of the differences between intelligibility scores on the first half compared to the second half which ranged from 0.04 to 0.38 percentage points. Both of these calculations suggested that none of the seven random divisions produced substantially greater or worse split-half reliability. Therefore, we chose the iteration that evenly distributed the number of words in each half (i.e., 55 words in each half; sentences 5, 6, 9, 10, 11, and 14 in half one and sentences 7, 8, 12, 13, and 15 in half two). The split-half

reliability (or intra-rater reliability) was 0.92. To ensure the same was true of speaking rate, a correlation between speaking rate on the first half of the SIT and speaking rate on the second half of the SIT was also computed for the chosen iteration. The intra-rater reliability for the speaking rate analysis was also 0.92 and the absolute mean of the differences between halves was 3.13 WPM, which demonstrated a strong association between outcomes on the first half of the SIT and outcomes on the second half of the SIT. The results of these analyses suggested an even division of the SIT sentences into two equal halves. The intra-rater reliability of 0.92 is also similar to the reliability of orthographic transcription found in previous work (Bunton, Kent, Kent, & Duffy, 2001; Hustad, 2006; Keintz, Bunton, & Hoit, 2007; Yorkston & Beukelman, 1978; 1981).

Riddle and Stratford (2013) identified two assumptions that must be met to calculate the MDC: "(1) patients' true values for the outcome of interest have not changed, and (2) the distribution of patients' difference scores between test and retest is consistent with a normal distribution" (p. 85). Because the entire SIT was administered on one occasion and then sentences were divided into random halves, patients' true values were unlikely to change between the two halves of the test. In addition, the difference of scores between the two halves of the SIT were checked for normality and, therefore, met the second assumption.

The MDC at the 95% confidence level was calculated based on standard estimation methods found in prior literature (Fulk & Echternach, 2008; Haley & Fragala-Pinkham, 2006; Kovacs et al., 2008; Lehman & Velozo, 2010; Mallinson, Pape, & Guernon, 2016; Riddle & Stratford, 2013) using the following formula: $MDC_{95} = 1.96 \times \sqrt{2} \times SEM$, where 1.96 is the z-score for the 95% confidence level, the square root of two accounts for the error in the two halves of the SIT, and the SEM is the standard error of the measurement. The MDC was

calculated for entire group of participants and separately for each stratum seen in Figure 1 for both sentence intelligibility and speaking rate.

Minimal clinically important difference. Only data from the participants with ALS was used for calculation of the MCID. Calculations of the MCID were modeled after Beninato et al. (2014) and Tilson et al. (2010). In order to estimate the MCID, we required two adjacent data collection sessions to calculate a change in score from the first session to the second session. The sample population was divided into participants who did or did not experience a "true" change in ALSFRS-R subscore (see Figure 2). Receiver Operating Characteristics (ROC) curves were used to determine how well the changes in sentence intelligibility and speaking rate differentiated between those patients who reported no change (no change on ALSFRS-R subscore) from those who reported change (a one or two-point change on the ALSFRS-R subscore). From ROC analyses, we obtained the optimal cut point which was defined as the MCID and maximized both sensitivity and specificity. The optimal cut point, or MCID, can also be estimated by visually determining the point on the ROC curve that is closest to the upper left-hand corner of the graph (Stratford & Riddle, 2012; Tilson et al., 2010). Sensitivity and specificity of the MCID in distinguishing between changed and unchanged patients was calculated. The area under the curve (AUC) was calculated to determine the probability that the test will be able to distinguish between patients who have changed from those who have not changed. The AUC can range from 0-1, with 0.50 indicating that the test is no better than chance at discriminating between patients. An AUC of 0.7 to 0.8 is considered to be adequate and an AUC of 0.8 to 0.9 is considered to be excellent (Copay, Subach, Glassman, Polly, & Schuler, 2007). MCIDs were calculated for the entire group of participants, first using a one-point change on the ALSFRS-R subscore as the cut-off score and then using a two-point change on the ALSFRS-R subscore as the cut-off score, and

separately for each stratum seen in Figure 2. R Development Core Team (2013) was used for statistical analysis.

Results

Results of the following analyses are presented in the following sections: overall MDC of sentence intelligibility and MDCs for each severity strata; overall MDC of speaking rate and MDCs for each severity strata; overall MCIDs for sentence intelligibility and speaking rate for a one-point change on the ALSFRS-R; overall MCIDs for sentence intelligibility and speaking rate for a two-point change on the ALSFRS-R; and MCIDs for sentence intelligibility and speaking rate for each severity strata.

Minimally Detectable Change

Sentence intelligibility. The overall MDC_{95} for sentence intelligibility, which was based on 639 observations, was a decline of 12.07%. MDCs for sentence intelligibility across the clinical stratification of severity groups are presented in the large graph of Figure 3. The MDC for the profound group (with 0-49% intelligibility; N=25) was a 24.82% decline in intelligibility, for the severe group (with 50-79% intelligibility; N=23) was a 41.12% decline in intelligibility, for the moderate group (with 80-89% intelligibility; N=25) was a 17.51% decline in intelligibility, for the mild group (with 90-95% intelligibility; N=25) was a 12.16% decline in intelligibility, and for the normal group (with 96-100% intelligibility, N=25) was a 3.55% decline in intelligibility. The MDCs of sentence intelligibility for the unbiased quantile stratification of dysarthria severity can be seen in the small panel in the top right hand corner of Figure 3. Visual inspection of the plots in Figure 3 suggest a similar ranking, of high to low, for both the clinical and the quantile stratifications.

[Figure 3]

Speaking rate. The overall MDC_{95} for speaking rate, which was based on 650 observations, was a decline of 36.57 WPM. MDCs for speaking rate across severity groups in the clinical stratification of severity are presented in the large graph of Figure 4. The MDC for the group with the fastest speaking rate (High SR; with >160 WPM; N=117) was a 40.86 WPM decline in speaking rate, for the group with mid-range speaking rate (Mid SR; with 120-160 WPM; N=143) was a 30.16 WPM decline in speaking rate, and for the group with the slowest speaking rate (Low SR; with <120 WPM; N=106) was a 22.45 WPM decline in speaking rate. The MDCs of speaking rate for the unbiased quantile stratification can be seen in the small panel in the top right hand corner of Figure 4. Again, visual inspection of the plots in Figure 4 suggest a similar ranking, of high to low, for both the clinical and the quantile stratifications.

[Figure 4]

Minimal Clinically Important Difference

Three hundred and sixty-six intervals (84 intervals with a decline of one point and 282 with no change) were included in the analyses for the MCID based on a one-point decline on the ALSFRS-R subscore. Overall, changes of 1.43% intelligibility (AUC = 0.63) and 8.80 WPM (AUC = 0.60) were identified from the ROC curve as the points that maximize sensitivity and specificity for dichotomizing participants into those that achieved the MCID threshold (i.e., one-point decline on the ALSFRS-R speech question between adjacent sessions) and those that did not achieve the MCID threshold (i.e., no change in ALSFRS-R subscore between adjacent sessions) (ROC curves can be seen in Figure 5).

[Figure 5]

Both AUCs indicate that the MCID values of 1.43% and 8.80 WPM may be no better than a 50/50 chance at distinguishing between those participants who reported having experienced

change (based on a one-point change in their ALSFRS-R subscore). Based on a two-point change on the ALSFRS-R subscore and 306 time intervals (24 intervals with a decline of two points and 282 intervals with no change), the AUC for sentence intelligibility was 0.83 (MCID = 3.18%) indicating that this two-point change did a better job of distinguishing between changed and unchanged patients. Based on a two-point change on the ALSFRS-R subscore for speaking rate the AUC was 0.68 (MCID = 8.44 WPM), which only minimally improved the ability of speaking rate to distinguish between changed and unchanged patients from a one-point change in ALSFRS-R subscore.

Results of the ROC analyses across severity groups are presented in Table 3. For participants who began with an ALSFRS-R subscore of four at the first session and changed to a subscore of three at an adjacent session, the MCID for intelligibility was 1.44% (AUC = 0.62) and the MCID for speaking rate was 19.7 WPM (AUC = 0.61). For participants who began with an ALSFRS-R subscore of three and changed to a subscore of two, the MCID for intelligibility was 4.55% (AUC = 0.67) and the MCID for speaking rate was 37.3 WPM (AUC = 0.58). Last, for participants who began with an ALSFRS-R subscore of two and changed to a subscore of one, the MCID for intelligibility was 1.78% (AUC = 0.71) and the MCID for speaking rate was 17.93 (AUC = 0.76).

[Table 3]

Discussion

The current study is, to our knowledge, the first to attempt to estimate the MDC and MCID of speech outcomes in a sizable clinical sample with a large range of sentence intelligibility and speaking rate deficits. MDC estimates for both outcomes were large relative to previously reported treatment effect sizes (see Table 1). These findings motivate the need for

additional research to determine if the intelligibility gains reported in speech intervention research are smaller than random measurement error. Furthermore, estimates of MCID lacked diagnostic accuracy, which impacts our current use of clinical outcome measures. The observed high MDC and MCID estimates may limit the interpretation of sentence intelligibility findings, and underscores the need for additional work focused on the development of reliable and psychometrically valid speech outcome measures.

The MDCs exceeded the size of most previously reported treatment effects

Overall, our estimates indicate that 95% of patients who have truly unchanged sentence intelligibility may display fluctuations in intelligibility up to 12%. Therefore, when a change occurs that is greater than our MDC of 12%, we can be confident that this change is real and is not due to random variability, at least for this population of patients with ALS as measured by the SIT. Across severities, these estimates vary widely, ranging from a 3.55% MDC in intelligibility for the normal group to a 41.12% MDC in intelligibility for the severe group. Comparing the percentage change values in intelligibility reported in the literature (Table 1) to our estimates of MDC reveals that we cannot be fully confident that many of the treatment gains in intelligibility are larger than random measurement error.

At the same time, the estimates of the MDC likely differ based on whether we are considering an increase or decrease in score (Coster, 2013; Hays & Wolley. 2000; Riddle & Stratford, 2013). Consequently, it is conceivable that an estimate of the MDC might be very different if improvements, rather than declines, in intelligibility are used. Further, the MDC is dependent on both the population and the context in which it is estimated, such that it is unlikely to obtain a single MDC estimate for a particular metric or population (Revicki et al., 2008). Therefore, the MDCs reported here may not hold true for other measures of intelligibility and

speaking rate, types of dysarthria with other etiologies, or even for other patients with ALS in different contexts. All of this considered, the estimates provided here are the first attempt to calculate MDCs of these constructs and should be considered a starting point for the exploration of these concepts in our fields.

Possible reasons for our high MDCs of intelligibility are related to the challenges associated with many intelligibility metrics. Measures of intelligibility are often highly variable, (Bunton et al., 2001; Hustad, 2006; Keintz et al., 2007; Yorkston & Beukelman, 1978; 1981) which lowers reliability and results in higher MDC estimates. A posthoc analysis of the current data revealed that intelligibility across the 11 SIT sentences was highly variable, with an average range of 19 percentage points and, in more severely affected subjects as high as 100 percentage points. There are several potential sources of this within-session variability in SIT scores including poor listener reliability, which, under the best conditions, is typically between .83 to .99 for similar tasks (Bunton et al., 2001; Hustad, 2006; Keintz et al., 2007; Yorkston & Beukelman, 1978; 1981). Another probable contributor to the high variation is differences among sentences in factors known to affect intelligibility such as phoneme distribution, phonotactic probability, word frequency, sonority (Kuruvilla-Dugdale, Green, Hogan, & Wang, 2017), and sentence length (Allison & Hustad, 2014). Ideally, the MDC would be estimated using reliability based on the same speech stimuli rather than a split-half approach. This approach, however, would lack ecological validity because it is inconsistent with how the SIT is scored. Future work should derive MDC estimates of intelligibility tests that reduce the possible variability in speaker and/or listener response, which may improve (or lower) calculations of the MDC. For example, an intelligibility test could reduce variability by having participants produce the same stimuli multiple times and have listeners score the same stimuli for each speaker.

However, this introduces confounding variables such as practice/familiarity effects for both speakers and listeners (D'Innocenzo, Tjaden, & Greenman, 2006; Tjaden & Liss, 1995). Improvement of intelligibility metrics is a challenging, but potentially very useful, endeavor.

The estimates of MDC for speaking rate in the current study indicate that, overall, a decline of 37 WPM is larger than measurement error. Across severities, speaking rate MDCs varied from 40.86 WPM for the group with fastest rate of speech to 22.45 WPM for the group with the slowest rate of speech. From a clinical perspective, measures of speaking rate serve two primary functions: (1) early detection of degenerative disease and (2) monitoring rate of disease progression. Green et al. (2013) examined group level differences in speaking rate between patients with ALS with (N = 10) and without (N = 10) intelligibility deficits and healthy controls (N = 10). To examine the potential of speaking rate for early disease detection, group level differences in speaking rate between controls and individuals with ALS without intelligibility deficits was approximately 30 WPM. Second, to examine the potential of speaking rate for monitoring rate of disease progression, group level differences in speaking rate between individuals with ALS without intelligibility deficits and individuals with ALS with intelligibility deficits was approximately 20 WPM. In the current study, the MDCs suggest that if patients with ALS have a slower speaking rate (i.e., likely further on in disease progression), we can be confident that small changes (i.e., close to the 20 WPM difference found by Green et al. [2013]) in speaking rate are real changes. In contrast, for patients with a faster speaking rate (i.e., likely earlier on in disease progression), we might not be as confident that a 30 WPM difference, that was reported by Green et al. (2013), is not due to measurement error.

Sentence intelligibility and speaking rate did not adequately distinguish between patients who reported change and those who did not report change on the ALSFRS-R subscore

Because of the large MDCs, we were unable to derive accurate estimates for minimal clinically important differences of intelligibility and speaking rate based on patients' ratings on the speech question of the ALSFRS-R. Our estimates for the MCIDs are smaller in magnitude than the estimates for the MDCs which is, theoretically, impossible (i.e., the smallest clinically relevant change cannot be smaller than the smallest detectable change). In past research, when the MDC has exceeded the value of the MCID, researchers have simply used the MDC instead (Riddle & Stratford, 2012). Similarly, most of the AUCs in the ROC analyses indicate an unacceptable level of diagnostic accuracy, further confirming the inability of sentence intelligibility and speaking rate in distinguishing between changed and unchanged patients, at least when change is solely defined by patients' scores on the ALSFRS-R speech question. A likely explanation is that there is some important and meaningful aspect of speech that influences patient self-ratings on the ALSFRS-R that is not accounted for by sentence intelligibility and speaking rate, such as the perceived amount of effort during speech. From this perspective, the one speech question on the ALSFRS-R is most likely an inadequate indicator of patient change. Future work could examine which aspects of speech production influence how patients perceive their own speech, which may inform other possible metrics that better capture important change for patients. Hays and Woolley (2000) highlight this as one problem with the external anchor process of calculating the MCID; that the estimate ultimately depends on the anchor chosen. Other possible measures with which to anchor the meaningfulness of changes in sentence intelligibility and speaking rate will be discussed in the future directions section.

When we raised the cut-off score on the ALSFRS-R for signaling important change to two points rather than one, sentence intelligibility did a better job of distinguishing between changed and unchanged patients (AUC = 0.83). Additionally, for the most severe group of

patients who began at an ALSFRS-R subscore of two and changed to a one, AUCs for both intelligibility and speaking rate were >0.70 with sensitivities of 1.00, suggesting that both measures were very sensitive to change in this group of patients. This finding suggests that even small changes in this most severe group is enough for patients to report that these changes are relevant. A second possibility is that patients are better able to report when they have changed from intelligible speech (i.e., score of two on the ALSFRS-R speech question) to speech combined with non-vocal communication (i.e., score of one on the ALSFRS-R speech question) than they are with reporting change on the upper half of the ALSFRS-R speech question (i.e., normal speech process [score of four] to detectable speech disturbance [score of three]).

MDC and MCID were dependent on severity

The results seen in Figures 3 and 4 suggest a similar ranking, of high to low, for both the clinical and the quantile stratifications. This suggests that the clinical stratification of participants did not have a significant influence on the calculations of the MDCs. Our estimates of the MDC and MCID differed widely as a function of dysarthria severity. For example, MDCs of sentence intelligibility ranged from 3.55% for the normal group to an MDC of 41.12% for the severe group. These findings highlight the importance of considering severity when calculating these types of estimates. Our findings also suggest that, when compared to groups with either profound or more mild intelligibility deficits, the group with mid-range intelligibility and/or self-perceived speech difficulty (i.e., ~50-79% intelligibility and/or patients who gave themselves an ALSFRS-R subscore of three), required a greater reduction in intelligibility both for this change to be considered ‘real’ and for patients to consider their change relevant or important. This finding is likely due to the greater variability in intelligibility in this more severely impaired group. Both speakers and listeners play an important role in this variability. Speakers with severe dysarthria

demonstrate variability in their production of speech (Blaney & Hewlett, 2007; Kuruvilla-Dugdale & Mefferd, 2017), and listeners have also reported decreased levels of confidence in transcribing dysarthric speech as severity of dysarthria increases (Hustad, 2007). Overall, both speakers with dysarthria and their communication partners struggle as dysarthria becomes more severe, particularly as intelligibility drops off significantly after 70% (Ball, Beukelman, & Pattee, 2004). Therefore, the importance of improving intelligibility, or expediting alternative means of communication, is accentuated in this group with severe deficits.

According to MDC estimates, for a change in speaking rate to be outside of measurement error, the change has to be greater for patients with a fast speaking rate than for patients with a slow speaking rate. Our MDC estimates ranged from 22.45 WPM for the group with the slowest speaking rate to 40.86 WPM for the group with the fastest speaking rate. The smaller range of speaking rate variability in the more affected group suggests that at this stage of the disease, muscle weakness has progressed to a level that limits the ability to accommodate the varying demands imposed by the speech stimulus.

Limitations and future directions

As previously mentioned, the ALSFRS-R might not be an adequate metric with which to anchor meaningfulness of change in sentence intelligibility or speaking rate. There is more work to be done on identifying anchors of meaningfulness for speech outcomes. Future studies could consider other anchors, such as comprehensibility (Yorkston et al., 1996), or other patient-reported measures such as the Communicative Participation Item Bank (CPIB; Baylor et al., 2009), Center for Neurologic Study Bulbar Function Scale (CNS-BFS; Smith et al., 2017), or a global rating of change scale like those often used in the physical therapy literature (Jaeschke et al., 1989). Stratford and Riddle (2012) stated: “For many important patient outcomes...it is

generally agreed that no error-free reference standard exists...We believe it is important for researchers to acknowledge the limitation of using any one of these methods and, where possible, to obtain threshold estimates using multiple reference standards..." (p. 1345). Future work could use a variety of referents and compare their ability to anchor change in these constructs.

Second, the MDC is customarily calculated using the test-retest reliability of a measure; however, in the current study, we used split-half reliability instead. This could have had some impact on our findings. Future studies could have speakers complete the SIT twice in order to calculate the test-retest reliability, as is typically done in studies estimating the MDC. However, we acknowledge that this would be challenging to do in very ill patients or in other patient groups whose status might fluctuate in between data collection sessions.

Finally, the current study explored the MDC and MCID only as a result of declines in sentence intelligibility and speaking rate. Future studies could look at these concepts for improvements in intelligibility and speaking rate as a result of speech-language therapy. Furthermore, future work could use this paradigm to estimate the MDC and MCID of a variety of speech outcomes for other populations in varying contexts.

Conclusions

In conclusion, we estimated that sentence intelligibility has an MDC of 12% and speaking rate has an MDC of 37 WPM for this group of patients with ALS as measured by the SIT**.** Although we were unable to derive usable estimates of clinically relevant changes in sentence intelligibility or speaking rate, there does appear to be value in exploring both the MDC and MCID as a function of baseline severity. In this study, we found that the mid-range severity group required a larger change in intelligibility for this change to be considered 'real' and 'important' when compared with the profound group and the less impaired groups. Overall,

application of the MDC and MCID to speech outcomes, such as sentence intelligibility and speaking rate, is an achievable and necessary endeavor.

References

Allison, K. M., & Hustad, K. C. (2014). Impact of sentence length and phonetic complexity on intelligibility of 5-year-old children with cerebral palsy. *International Journal of Speech-Language Pathology, 16*(4), 396-407.

Ball, L., Beukelman, D. R., & Pattee, G. (2002). Timing of speech deterioration in people with amyotrophic lateral sclerosis. *Journal of Medical Speech-Language Pathology, 10*, 231-235.

Baylor, C. R., Yorkston, K. M., Eadie, T. L., Miller, R. M., & Amtmann, D. (2009). Developing the communicative participation item bank: Rasch analysis results from a spasmodic dysphonia sample. *Journal of Speech, Language, and Hearing Research, 52*(5), 1302-1320.

Beckerman, H., Roebroeck, M. E., Lankhorst, G. J., Becher, J. G., Bezemer, P. D., & Verbeek, A. L. M. (2011). Smallest real difference, a link between reproducibility and responsiveness. *Quality of Life Research, 10*, 571-578.

Beninato, M., Fernandes, A., & Plummer, L. S. (2014). Minimal clinically important difference of the functional gait assessment in older adults. *Physical, 94*(11), 1594-1603.

Beninato, M., & Portney, L. G. (2011). Applying concepts of responsiveness to patient management in neurologic physical therapy. *Journal of Neurologic Physical Therapy, 35*, 75-81.

Beninato, M., Gill-Body, K. M., Salles, S., Stark, P. C., Black-Schaffer, R. M., & Stein, J. (2006). Determination of the minimal clinically important difference in the FIM instrument in patients with stroke. *Archives of Physical Medicine and Rehabilitation, 87*(1), 32-39.

Blaney, B., & Hewlett, N. (2007). Dysarthria and Friedreich's ataxia: What can intelligibility assessment tell us? *International Journal of Language & Communication Disorders, 42*(1), 19-37.

Boliek, C. A., & Fox, C. M. (2014). Individual and environmental contributions to treatment outcomes following a neuroplasticity-principled speech treatment (LSVT: LOUD) in children with dysarthria secondary to cerebral palsy: A case study review. *International Journal of Speech-Language Pathology, 16*(4), 372-385.

Brooks, B. R., Miller, R. G., Swash, M., & Munsat, T. L. (2000). El Escorial revisited: Revised criteria for the diagnosis of amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis and other Motor Neuron Disorders, 1*(5), 293-299.

Bunton, K., Kent, R. D., Kent, J. F., & Duffy, J. R. (2001). The effects of flattening fundamental frequency contours on sentence intelligibility in speakers with dysarthria. *Clinical Linguistics & Phonetics, 15*(3), 181-193.

Cannito, M. P., Suiter, D. M., Beverly, D., Chorna, L., Wolf, T., & Pfeiffer, R. M. (2012). Sentence intelligibility before and after voice treatment in speakers with idiopathic Parkinson's disease. *Journal of Voice, 26*(2), 214-219.

Castrillo-Viguera, C., Grasso, D. L., Simpson, E., Shefner, J., & Cudkowicz, M. E. (2010). Clinical significance in the change of decline in ALSFRS-R. *Amyotrophic Lateral Sclerosis, 11*, 178-180.

Cedarbaum, J. M., Stambler, N., Malta, E., Fuller, C., Hilt, D., Thurmond, B.,...BDNF ALS Study Group. (1999). The ALSFRS-R: A revised functional rating scale that incorporates assessments of respiratory function. *Journal of the Neurological Sciences, 169*, 13-21.

Connaghan, K. P., & Patel, R. (2017). The impact of contrastive stress on vowel acoustics and

intelligibility in dysarthria. *Journal of Speech, Language, and Hearing Research, 60*, 38-50.

Copay, A. G., Subach, B. R., Glassman, S. D., Polly, D. W., & Schuler, T. C. (2007). Understanding the minimum clinically important difference: A review of concepts and methods. *The Spine Journal, 7*, 541-546.

Coster, W. J. (2013). Making the best match: Selecting outcome measures for clinical trials and outcome studies. *The American Journal of Occupational Therapy, 67*(2), 162-170.

Dagenais, P. A., Brown, G. R., & Moore, R. E. (2006). Speech rate effects upon intelligibility and acceptability of dysarthric speech. *Clinical Linguistics & Phonetics, 20*(2-3), 141-148.

De Vet, H. C., Terwee, C. B., Ostelo, R. W., Beckerman, H., Knol, D. L., & Bouter, L. M. (2006). Minimal changes in health status questionnaires: Distinction between minimally detectable change and minimally important change. *Health and Quality of Life Outcomes, 4*(54), 1-5.

D'Innocenzo, J., Tjaden, K., & Greenman, G. (2006). Intelligibility in dysarthria: Effects of listener familiarity and speaking condition. *Clinical Linguistics & Phonetics, 20*(9), 659-675.

Donovan, N. J., Kendall, D. L., Young, M. E., & Rosenbek, J. C. (2008). The communicative effectiveness survey: Preliminary evidence of construct validity. *American Journal of Speech-Language Pathology, 17*(4), 335-347.

dos Santos Barreto, S., & Zazo Ortiz, K. (2016). Protocol for the evaluation of speech intelligibility in dysarthrias: Evidence of reliability and validity. *Folia Phoniatrica et Logopaedica, 67*(4), 212-218.

Doyle, P. C., Leeper, H. A., Kotler, A., Thomas-Stonell, N., O'Neill, C., Dylke, M., & Rolls, K. (1997). Dysarthric speech: A comparison of computerized speech recognition and listener intelligibility. *Journal of Rehabilitation Research and Development, 34*(3), 309-316.

Fulk, G. D., & Echternach, J. L. (2008). Test-retest reliability and minimal detectable change of gait speed in individuals undergoing rehabilitation after stroke. *Journal of Neurologic Physical Therapy, 32*(1), 8013.

Green, J. R., Yunusova, Y., Kuruvilla, M. S., Wang, J., Pattee, G. L., Synhorst, L., Zinman, L., & Berry, J. D. (2013). Bulbar and speech motor assessment in ALS: Challenges and future directions. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration, 14*(7-8), 494-500.

Gordon, P. H., Cheng, B., Montes, J., Doorish, C., Albert, S. M., & Mitsumoto, H. (2007). Outcome measures for early phase clinical trials. *Amyotrophic Lateral Sclerosis, 8*2, 270-273.

Haley, S. M., & Fragala-Pinkham, M. A. (2006). Interpreting change scores of tests and measures in physical therapy. *Physical Therapy, 86*(5), 735-743.

Hammen, V. L., Yorkston, K. M., & Minifie, F. D. (1994). Effects of temporal alterations on speech intelligibility in parkinsonian dysarthria. *Journal of Speech and Hearing Research, 37*, 244-253.

Hartelius, L., Elmberg, M., Holm, R., Lövberg, A., & Nikolaidis, S. (2008). Living with dysarthria: Evaluation of a self-report questionnaire. *Folia Phoniatrica et Logopaedica, 60*, 11-19.

Hays, R. D., & Woolley, J. M. (2000). The concept of clinically meaningful difference in health-related quality-of-life research: How meaningful is it? *Pharmacoeconomics, 18*(5), 419-423.

Hustad, K. C. (2008). The relationship between listener comprehension and intelligibility scores for speakers with dysarthria. *Journal of Speech, Language, and Hearing Research, 51,* 562-573.

Hustad, K. C. (2007). Effects of speech stimuli and dysarthria severity on intelligibility scores and listener confidence ratings for speakers with cerebral palsy. *Folia Phoniatrica et Logopaedica, 59*(6), 306-317.

Hustad, K. C. (2006). A closer look at transcription intelligibility for speakers with dysarthria: Evaluation of scoring paradigms and linguistic errors made by listeners. *American Journal of Speech-Language Pathology, 15*(3), 268-277.

Hustad, K. C., Oakes, A., & Allison, K. (2015). Variability and diagnostic accuracy of speech intelligibility scores in children. *Journal of Speech, Language, and Hearing Research, 58*, 1695-1707.

Jaeschke, R., Singer, J., & Guyatt, G. H. (1989). Measurement of health status: Ascertaining the minimal clinically important difference. *Controlled Clinical Trials, 10*(4), 407-415.

Kaufmann, P., Levy, G., Montes, J., Buchsbaum, R., Barsdorf, A. I., Battista, V.,...Thompson, J. L. P. (2007). Excellent inter-rater, intra-rater, and telephone-administered reliability of the ALSFRS-R in a multicenter clinical trial. *Amyotrophic Lateral Sclerosis, 8*, 42-46.

Keintz, C. K., Bunton, K., & Hoit, J. D. (2007). Influence of visual information on the intelligibility of dysarthric speech. *American Journal of Speech-Language Pathology, 16*, 222-234.

Kent, R. D., Weismer, G., Kent, J. F., & Rosenbek, J. C. (1989). Toward phonetic intelligibility testing in dysarthria. *The Journal of Speech and Hearing Disorders, 54*(4), 482-499.

Kim, H., Martin, K., Hasegawa-Johnson, M., & Perlman, A. (2010). Frequency of consonant articulation errors in dysarthric speech. *Clinical Linguistics & Phonetics, 24*(10), 759-770.

Kimura, F., Fujimura, C., Ishida, S., Nakajima, H., Furutama, D., Uehara, H.,…Hanafusa, T. (2006). Progression rate of ALSFRS-R at time of diagnosis predicts survival time in ALS. *Neurology, 66*, 265-267.

Kollewe, K., Mauss, U., Krampfl, K., Petri, S., Dengler, R., & Mohammdi, B. (2008). ALSFRS-R score and its ratio: A useful predictor for ALS-progression. *Journal of the Neurological Sciences, 275*, 69-73.

Kovacs, F. M., Abraira, V., Royuela, A., Corcoll, J., Algere, L., Tomas, M.,...Mufraggi, N. (2008). Minimum detectable and minimal clinically important changes for pain in patients with nonspecific neck pain. *BioMed Central Musculoskeletal Disorders, 9*(43), 1-9.

Kuruvilla-Dugdale, M. S., Green, J. R., Hogan, T. P., & Wang, J. (2017). Phonological complexity effects on tongue motor control. Manuscript submitted for publication.

Kuruvilla-Dugdale, M., & Mefferd, A. (2017). Spatiotemporal movement variability in ALS: Speaking rate effects on tongue, lower lip, and jaw motor control. *Journal of Communication Disorders, 67*, 22-34.

Lam, J., & Tjaden, K. (2013). Intelligibility of clear speech: Effect of instruction. *Journal of Speech, Language, and Hearing Research, 56*, 1429-1440.

Lehman, L. A., & Velozo, C. A. (2010). Ability to detect change in patient function: Responsiveness designs and methods of calculation. *Journal of Hand Therapy, 23*(4), 361-371.

Mallinson, T., Pape, T. L., & Guernon, A. (2016). Responsiveness, minimal detectable change, and minimally clinically important differences for the disorders of consciousness scale. *The Journal of Head Trauma Rehabilitation, 31*(4), 43-51.

McAuliffe, M. J., Fletcher, A. R., Kerr, S. E., O'Beirne, G. A., & Anderson, T. (2017). Effect of dysarthria type, speaking condition, and listener age on speech intelligibility. *American Journal of Speech-Language Pathology, 26,* 113-123.

Miller, N. (2013). Review: Measuring up to speech intelligibility. *International Journal of Language & Communication Disorders, 48*(6), 601-612.

Nouri, F. M., & Lincoln, N. B. (1987). An extended activity of daily living scale for stroke patients. *Clinical Rehabilitation, 1,* 301-305.

Nunnally J. C., & Bernstein, I. H. (1994). *Psychometric theory.* New York: McGraw- Hill.

Paganoni, S., Cudkowicz, M., & Berry, J. D. (2013). Outcome measures in amyotrophic lateral sclerosis clinical trials. *Clinical Investigation, 4*(7), 605-618.

Park, S., Theodoros, D., Finch, E., & Cardell, E. (2016). Be clear: A new intensive speech treatment for adults with nonprogressive dysarthria. *American Journal of Speech-Language Pathology, 25,* 97-110.

Pilon, M. A., McIntosh, K. W., & Thaut, M. H. (1998). Auditory vs visual speech timing cues as external rate control to enhance verbal intelligibility in mixed spastic ataxic dysarthric speakers: A pilot study. *Brain Injury, 12*(9), 793-803.

Plowman, E. K., Watts, S. A., Tabor, L., Robison, R., Gaziano, J., Domer, A. S.,…Gooch, C.

(2016). Impact of expiratory strength training in amyotrophic lateral sclerosis. *Muscle & Nerve, 53*(1), 48-53.

R Development Core Team. (2013). R: A language and environment for statistical computing [Software]. R Foundation for Statistical Computing, Vienna: Austria. Available from https://www.r-project.org/.

Revicki, D., Hays, R. D., Cella, D., & Sloan, J. (2008). Recommended methods for determining responsiveness and minimally important difference for patient-reported outcomes. *Journal of Clinical Epidemiology, 61*(2), 102-109.

Riddle, D., & Stratford, P. (2013). *Is this change real? Interpreting patient outcomes in physical therapy.* Philadelphia, PA: Davis Plus.

Rong, P., Yunusova, Y., Wang, J., & Green, J. R. (2015). Predicting early bulbar decline in amyotrophic lateral sclerosis: A speech subsystem approach. *Behavioural Neurology,* 1-11.

Rong, P., Yunusova, Y., Wang, J., Zinman, L., Pattee, G. L., Berry, J. D., … & Green, J. R. (2016). Predicting speech intelligibility decline in amyotrophic lateral sclerosis based on the deterioration of the individual speech subsystems. *Plos One, 22*(5), 1-19.

Shellikeri, S., Green, J. R., Kulkarni, M., Rong, P., Martino, R., Zinman, L., & Yunusova, Y. (2016). Speech movement measures as markers of bulbar disease in amyotrophic lateral sclerosis. *Journal of Speech, Language, and Hearing Research, 25,* 1-13.

Smith, R., Pioro, E., Myers, K., Sirdofsky, M., Goslin, K., Meekins, G.,…Pattee, G. (2017). Enhanced bulbar function in amyotrophic lateral sclerosis: The nuedexta treatment trial. *Neurotherapeutics, 14*(3), 762-772.

Stipancic, K. L., Tjaden, K., & Wilding, G. (2016). Comparison of intelligibility measures for adults with Parkinson's disease, adults with multiple sclerosis, and healthy controls. *Journal of Speech, Language, and Hearing Research, 59*(2), 1-9.

Stratford, P., & Riddle, D. (2012). When minimal detectable change exceeds a diagnostic test based threshold change value for an outcome measure: Resolving the conflict. *Physical Therapy, 92*(1), 1338-1347.

Streiner, D. L., & Norman, G. R. (2008). *Health measurement scales: A practical guide to their development and use* (4th ed.). New York: Oxford University Press.

Tilson, J. K., Sullivan, K. J., Cen, S. Y., Rose, D. K., Koradia, C. H., Azen, S. P., & Duncan, P. W. (2010). Meaningful gait speed improvement during the first 60 days poststroke: Minimal clinically important difference. *Physical Therapy, 90*(2), 196-208.

Tjaden, K., Kain, A., & Lam, J. (2014). Hybridizing conversational and clear speech to investigate the source of increased intelligibility in speakers with Parkinson's disease. *Journal of Speech, Language, and Hearing Research, 57*, 1191-1205.

Tjaden, K. K., & Liss, J. M. (1995). The role of listener familiarity in the perception of dysarthric speech. *Clinical Linguistics & Phonetics, 9*(2), 139-154.

Traynor, B. J., Zhang, H., Shefner, J. M., Schoenfeld, D., Cudkowicz, M. E., & on behalf of the NEALS Consortium (2004). Functional outcome measures as clinical trial endpoints in ALS. *Neurology, 63*(10), 1933-1935.

Van Nuffelen, G., De Bodt, M., Vanderwegen, J., Van de Heyning, P., & Wuyts, F. (2010). Effect of rate control on speech production and intelligibility in dysarthria. *Folia Phoniatrica et Logopaedica, 62*, 110-119.

Walshe, M., Peach, R., & Miller, N. (2009). Dysarthria impact profile: Development of a scale to measure psychosocial effects. *International Journal of Language and Communication Disorders, 44*(5), 693-715.

Wu, C., Chuang, L., Lin, K., Lee, S., & Hong, W. (2011). Responsiveness, minimal detectable change, and minimal clinically important difference of the Nottingham extended activities of daily living scale in patients with improved performance after stroke rehabilitation. *Archives of Physical Medicine and Rehabilitation, 92*(8), 1281-1287.

Wang, Y., Hart, D. L., Stratford, P. W., & Mioduski, J. E. (2011). Baseline dependency of minimal clinically important improvement. *Physical Therapy, 91*(5), 675-688.

Yorkston, K., & Beukelman, D. (1981). Communication efficiency of dysarthric speakers as measured by sentence intelligibility and speaking rate. *Journal of Speech and Hearing Disorders, 46*(3), 296-301.

Yorkston, K., & Beukelman, D. (1978). A comparison of techniques for measuring intelligibility of dysarthric speech. *Journal of Communication Disorders, 11*(6), 599-512.

Yorkston, K., Beukelman, D., & Hakel, M. (2007). Speech Intelligibility Test (SIT) for Windows [computer software]. Lincoln, Nebraska, USA: Madonna Rehabilitation Hospital.

Yorkston, K. M., Hammen, V. L., Beukelman, D. R., & Traynor, C. D. (1990). The effect of rate control on the intelligibility and naturalness of dysarthric speech. *Journal of Speech and Hearing Disorders, 55*(3), 550-560.

Yunusova, Y., Green, J. R., Lindstrom, M. J., Ball, L. J., Pattee, G. L., & Zinman, L. (2010). Kinematics of disease progression in bulbar ALS. *Journal of Communication Disorders, 43*, 6-20.

Yunusova, Y., Green, J. R., Wang, J., Pattee, G. & Zinman, L. (2011). A protocol for

comprehensive assessment of bulbar dysfunction in amyotrophic lateral sclerosis (ALS). *Journal of Visualized Experiments*, *48*, 1-4.

Table 1. Gains in intelligibility in dysarthria literature as a result of speech-language therapy techniques.

Authors (Date)	Population	Therapy Technique	Measure of Intelligibility	Improvement in Intelligibility
Boliek & Fox (2014)	2 children with CP	Lee Silverman Voice Treatment (LSVT)	Transcription	One child with 11% increase in intelligibility; one child with 16% increase in intelligibility
Dagenais, Brown, & Moore (2006)	4 adults with dysarthria (etiology: strokes, ALS) and 2 control speakers	Fast speech and slow speech	Transcription	No improvements in intelligibility
D'Innocenzo, Tjaden, & Greenman (2006)	1 adult with dysarthria secondary to TBI	Fast speech and loud speech	Transcription	Fast speech vs. habitual: 3% increase in intelligibility Loud speech vs. habitual: 24% increase in intelligibility
Hammen, Yorkston, & Minifie (1994)	5 adults with Parkinson's disease	Slow, paced speech	Transcription	Slow condition vs. habitual: 2% decrease-13% increase in intelligibility
McAuliffe, Fletcher, Kerr, O'Beirne, & Anderson (2017)	6 adults with dysarthria (etiology not specified; 3 with ataxic dysarthria and 3 with hypokinetic dysarthria)	Loud and slow speech	Listeners verbally repeated phrases that were heard; listener repetitions were transcribed	Loud condition vs. habitual: 12-17% increase in intelligibility Slow condition vs. habitual: 4-19% increase in intelligibility
***Park, Theodoros, Finch, & Cardell (2016)**	8 adults with dysarthria (6 TBI, 2 stroke)	Be Clear program; clear speech	Transcription	On average: 8% increase in intelligibility
Pilon, McIntosh, & Thaut (1998)	3 adults with dysarthria secondary to TBI	Pacing strategies	Transcription	Pacing strategies vs. habitual: 8% decrease to 20% increase in intelligibility

Yorkston, Hammen, Beukelman, & Traynor (1990)	8 adults with dysarthria (3 with PD, 1 with CP, 1 with cerebellar degeneration, 2 with TBI, and 1 with a tumor resection)	Slow speech	Transcription	Slow conditions vs. habitual: 21-33% increase in intelligibility
***^Cannito, Suiter, Beverly, Chorna, Wolf, & Pfeiffer (2012)**	8 adults with Parkinson's disease	Lee Silverman Voice Treatment (LSVT)	Transcription (stimuli presented in the presence of pink noise)	On average: 4% increase in intelligibility
^Lam & Tjaden (2013)	12 healthy adults	Clear speech (conditions with different instructions for production)	Transcription (stimuli mixed with multi-talker babble)	Clear speech vs. habitual: 16%-28% increase in intelligibility
^Stipancic, Tjaden, & Wilding (2016)	32 healthy adults, 16 adults with dysarthria secondary to PD, 30 adults with dysarthria secondary to MS	Clear, loud, and slow speech	Transcription (stimuli mixed with multi-talker babble)	On average: Clear condition vs. habitual: 5%-12% increase in intelligibility Loud condition vs. habitual: 5%-11% increase in intelligibility Slow condition vs. habitual: 3% decrease-4% increase in intelligibility

NOTE: * indicates treatment studies; all others were stimulability studies. ^ indicates studies that presented stimuli mixed with background noise – these studies should cautiously be compared to those in which stimuli were presented in quiet. Of note, only studies that used transcription to measure intelligibility have been included in this table to allow for comparison of reported improvements in intelligibility to the findings in the current paper.

Table 2. Validation of list equivalency for split halves approach on the SIT.

List 1 Sentences	List 1 # of words	List 2 Sentences	List 2 # of words	r of intelligibility	Mean of intelligibility differences
*5, 6, 9, 10, 11, 14	55	7, 8, 12, 13, 15	55	0.92	0.35
5, 7, 9, 10, 12, 15	58	6, 8, 11, 13, 14	52	0.95	-0.04
5, 7, 10, 11, 13, 15	60	6, 8, 10, 12, 14	50	0.94	-0.2
5, 8, 9, 12, 15	49	6, 7, 10, 11, 13, 14	61	0.92	-0.28
5, 8, 9, 10, 13, 14	59	6, 7, 11, 12, 15	51	0.94	0.11
5, 6, 9, 11, 13, 15	59	7, 8, 10, 12, 14	51	0.93	-0.38
^5, 7, 9, 11, 13	45	6, 8, 10, 12, 14	50	0.93	0.2

For each iteration of the split halves, the 11 sentences of the SIT were randomly divided into two lists. In each iteration, we calculated the total number of words in each list, the correlation (r) between the two halves, and the mean of the differences between the halves. NOTE: Because no iteration substantially improved the correlation between the two halves, and the mean of the differences was relatively normally distributed (as visually examined on histograms) for each iteration, * indicates the division that was chosen because it equally distributed the number of words in each half. ^ indicates that sentence #15 (longest sentence length) was removed in this iteration to examine whether this would change the correlation between halves.

Table 3. Minimal clinically important differences (MCID) of intelligibility and speaking rate, cutoff scores, sensitivity, and specificity. 95% confidence intervals (CI) are presented in parentheses.

Cut-off score on ALSFRS-R subscore	MCID	AUC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Intelligibility				
One point change	1.44%	0.63 (0.56-0.70)	0.48 (0.36-0.58)	0.77 (0.72-0.82)
Two point change	3.18%	0.83 (0.75-0.92)	0.67 (0.46-0.83)	0.87 (0.83-0.91)
Score of 4 to score of 3	1.44%	0.62 (0.53-0.70)	0.41 (0.28-0.52)	0.79 (0.74-0.85)
Score of 3 to score of 2	4.55%	0.67 (0.51-0.84)	0.56 (0.33-0.78)	0.84 (0.75-0.91)
Score of 2 to score of 1	1.78%	0.71 (0.45-0.97)	1 (1.00-1.00)	0.57 (0.28-0.86)
Speaking Rate				
One point change	8.80 WPM	0.60 (0.53-0.67)	0.70 (0.61-0.80)	0.52 (0.46-0.58)
Two point change	8.44 WPM	0.68 (0.58-0.79)	0.88 (0.75-1.00)	0.51 (0.46-0.57)
Score of 4 to score of 3	19.7 WPM	0.61 (0.53-0.70)	0.54 (0.43-0.67)	0.72 (0.66-0.79)
Score of 3 to score of 2	37.3 WPM	0.58 (0.42-0.74)	0.50 (0.28-0.72)	0.75 (0.65-0.84)
Score of 2 to score of 1	17.93 WPM	0.76 (0.53-0.99)	1 (1.00-1.00)	0.71 (0.50-0.93)

The cut-off score on the ALSFRS-R speech question was the score used to distinguish between patients who have changed and patients who have not changed. Responses on the speech question range from 4 (normal speech) to 0 (loss of useful speech). The MCIDs are relatively small, although the group who began with a score of 3 and changed to a score of 2 had the highest MCID for both intelligibility and speaking rate. The diagnostic accuracy represented by most of the AUCs, sensitivity values, and specificity values is low, except for when a two point change on the ALSFRS-R subscore was used as a cut-off score.

Figures & Captions

Figure 1. Schematic of Minimally Detectable Change (MDC) analyses. NOTE: N for MDC_{Total} reflects the total number of subjects in the sample with available data. The number of subjects in the subgroup analyses has been reduced to allow for a similar number of participants in each subgroup.

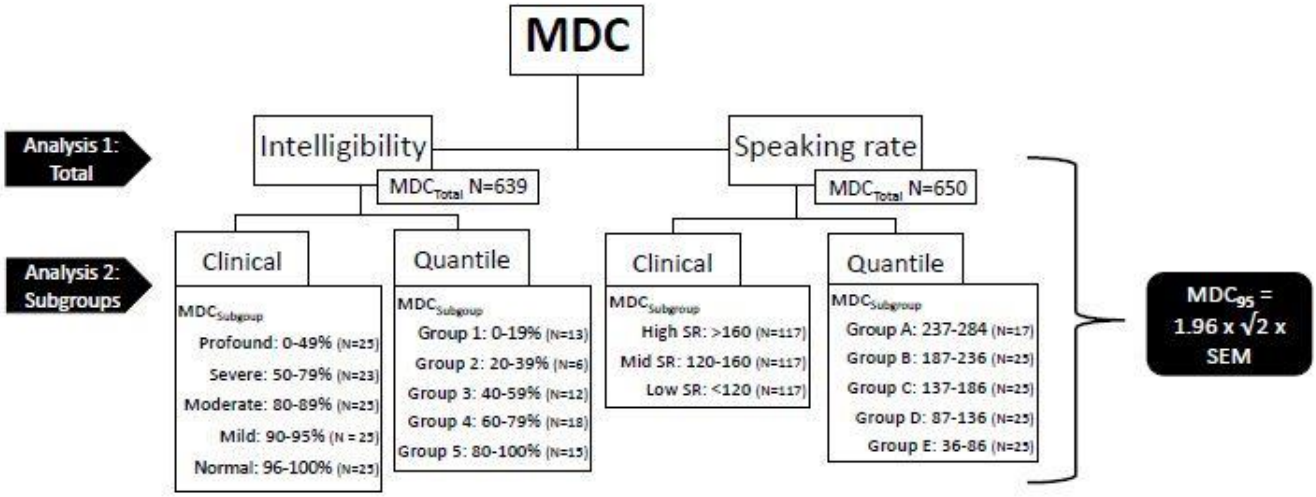

Figure 2. Schematic of Minimal Clinically Important Difference (MCID) analyses. ALSFRS-R Speech question subscore: "How is your speech?" Possible responses: (4) Normal speech process, (3) Detectable speech disturbance, (2) Intelligible with repeating, (1) Speech combined with non-vocal communication, (0) Loss of useful speech. NOTE: In each analysis, the participants who reached the threshold of change on the ALSFRS-R subscore were compared to the participants who did not reach the threshold of change on the ALSFRS-R subscore (i.e., "no change" or "stayed at score of __").

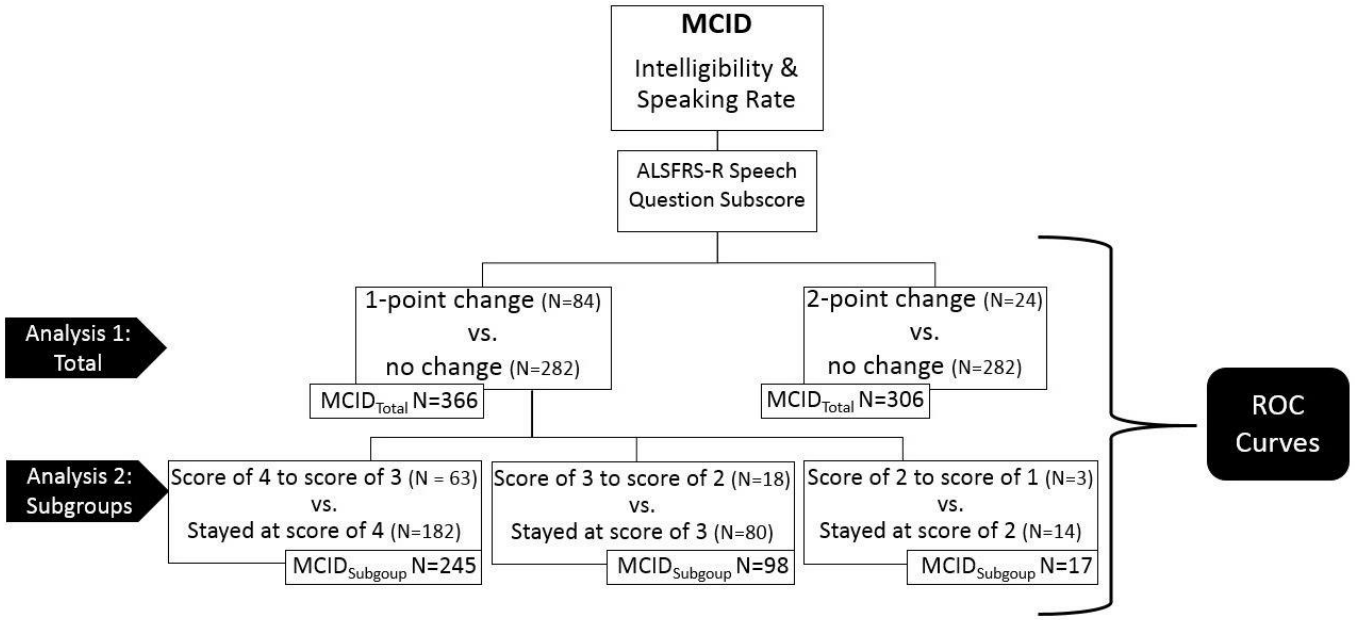

Figure 3. MDCs of intelligibility across severity levels based on a clinical stratification of dysarthria severity (large graph) and based on a quantile stratification of dysarthria severity (small graph). In both stratifications, the severe group has the highest MDC for intelligibility and the normal group has the lowest MDC for intelligibility. Percent intelligibility for each group on the graph is as follows: profound = 0-49%, severe = 50-70%, moderate = 80-89%, mild = 90-95%, normal = 96-100%, group 1 = 0-19%, group 2 = 20-39%, group 3 = 40-59%, group 4 = 60-79%, group 5 = 80-100%.

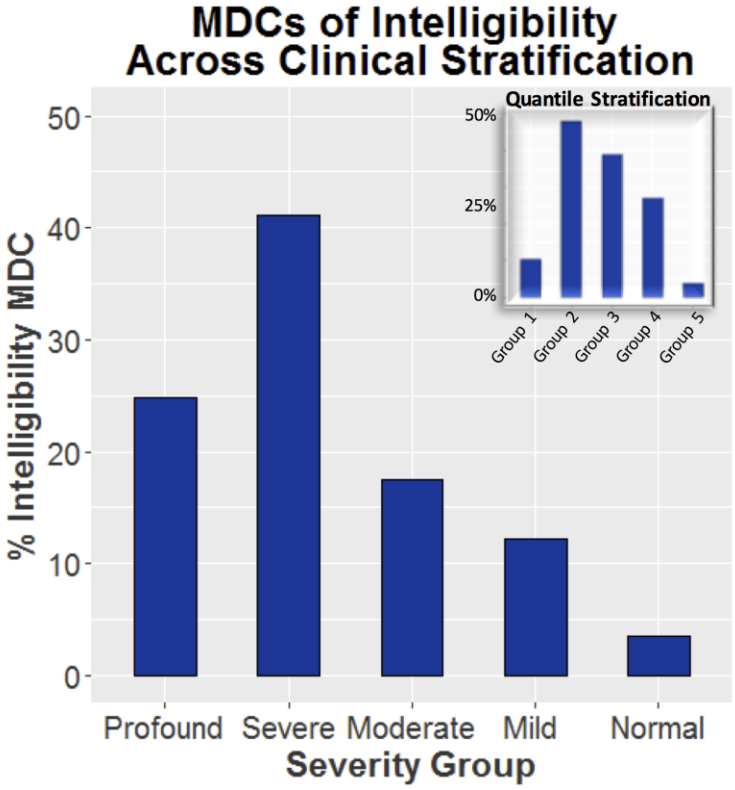

Figure 4. MDCs of speaking rate across severity levels based on a clinical stratification of dysarthria severity (large graph) and based on a quantile stratification of dysarthria severity (small graph). In both stratifications, the groups with the highest speaking rates have the highest MDCs for speaking rate and groups with the slowest speaking rates have the lowest MDCs for speaking rate. Speaking rate for each group on the graph is as follows: High SR = >160 WPM, Mid SR = 120-160 WPM, Low SR = <120 WPM, group A = 237-284 WPM, group B = 187-236 WPM, group C = 137-186 WPM, group D = 87-136 WPM, group E = 36-86 WPM.

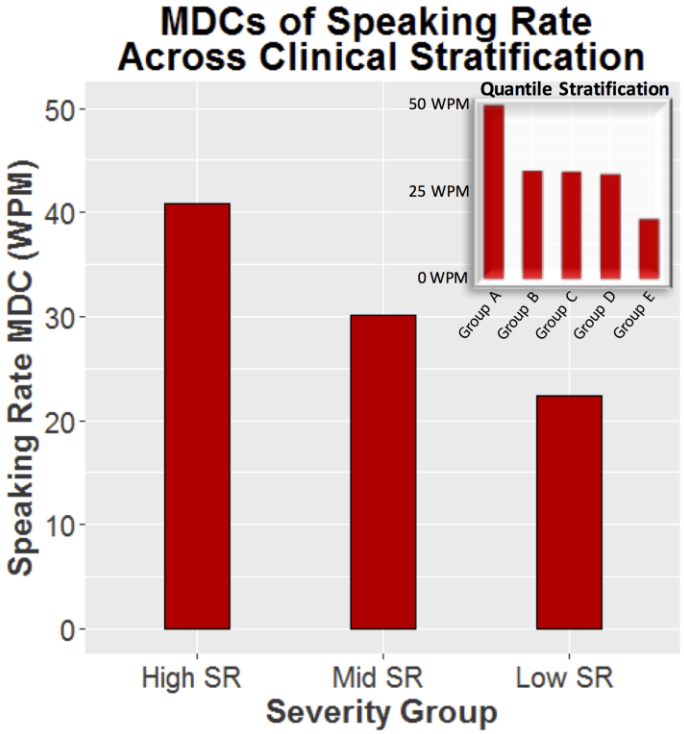

Figure 5. Receiver Operating Characteristic (ROC) curves for a one-point change on the ALSFRS-R subscore and a two-point change on the ALSFRS-R subscore. The MCID is the point closest to the upper left hand corner of the curves and represents the point that maximizes sensitivity and specificity. The straight black line represents no better than chance of distinguishing between changed and unchanged participants. Using a two-point change on the ALSFRS-R subscore as the threshold for true change improves the diagnostic accuracy of both sentence intelligibility and speaking rate in distinguishing between truly changed and unchanged participants.