Identifying Elevated Rates of CDR Scoring Errors: The Cognitive-Functional Difference Score

Christopher Weber[1], Selam Negash[1], Michael T. Ropacki[1], Christopher Randolph[1,2]

[1] MedAvante, Inc. [2] Loyola University Medical Center

BACKGROUND

- The Clinical Dementia Rating scale (CDR) is a subjectively rated clinician-reported outcome measure, used as a sole primary endpoint in clinical trials of early symptomatic Alzheimer's disease (AD).
- The CDR is also a key secondary endpoint that contributes to diagnostic classification in prevention trials for time-to event analyses, and also used as a co-primary endpoint in some trials of mild to moderate dementia due to AD[1].
- Scoring the CDR can be challenging, particularly in mild disease; rater administration and scoring errors are common.
- The goal of this study was to identify a CDR quality metric, based upon internal patterns of domain scores that could be used to identify atypical patterns that might be indicative of scoring errors.

METHOD

- 10,283 CDR assessments, with CDR sum of boxes (CDR-SB) ranging from 0.5 to 5, from eleven ongoing industry-sponsored clinical trials in Alzheimer's disease, including prevention, prodromal, and mild-moderate dementia were collated for analyses.
- The CDR-SB range of 0.5 to 5 was chosen to restrict the focus of the study to mild disease, where the scoring is typically more challenging[2].
- All CDR assessments were independently reviewed by trained and calibrated clinicians. A preliminary pattern analysis revealed an expected pattern of elevations in scores (more impairment) on the three cognitive domains (Memory, Orientation, Judgment & Problem-Solving) that preceded and remained above elevations in the three functional domains (Community Affairs, Home & Hobbies, and Personal Care) with increasing CDR-SB scores.
- We decided therefore to construct a single score that would reflect the relationship of the cognitive domains to the functional domains by taking the mean of the cognitive domains for each CDR assessment and subtracting the mean of the functional domains.
- Our hypothesis was that this score, referred to as the cognitive-functional difference (CFD) score should be positive in the vast majority of cases, and that negative scores might be associated with increased scoring errors.
- To test this hypothesis, the CFD scores for each assessment were calculated from site rater scores, and errors as detected by independent review were examined for both "typical" (positive) CFD scores, and "atypical" (0 or negative) CFD scores.

RESULTS

- For the overall sample, 89.2 percent of the CFD scores were positive (N=9,167), and 10.8 percent were at or below zero.
- In order to further explore the pattern of positive and negative CFD scores as a function of overall CDR-SB, assessments were clustered into CDR-SB categories of 1,2,3,4, and 5, with half-step scores being included in the next higher full-step score (e.g., a CDR-SB of 3.5 would be included with CDR-SBs of 4).
- Figure 1 shows the frequency distributions by CDR-SB categories of "typical" (positive) CFD scores, and "atypical" (0 or negative) CFD scores.
- The percentage of positive CFD scores across these 5 steps ranged from 85 to 92 percent, with no discernible pattern of change as a function of overall CDR-SB.
- Figure 2 shows the percentage of reviews with errors for typical and atypical groups as a function of CDR-SBs. Error rates were significantly higher in the atypical group compared to typical, $F(1,10,281) = 142.3$, $p < .0001$. At each CDR-SB level, the atypical group had significantly higher errors compared to typical, (p-values <.0001).
- Comparison of clinician reviewer scores revealed that in atypical CFD scores with at least one error (n=181), clinician reviewers improved (i.e., shifted CFD closer to a typical score) in 94 percent of the assessments (Figure 3).

Figure 1: Distribution of Typical and Atypical Scores by CDR-SB

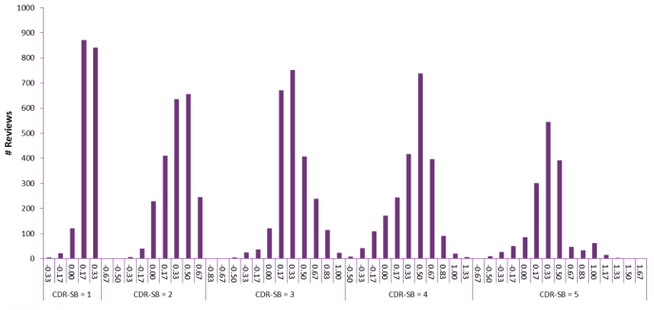


Figure 2: Percent Reviews with Discrepancies for Typical and Atypical Scores

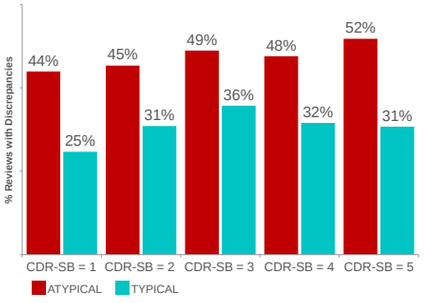
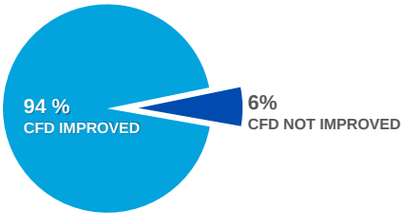


Figure 3: Atypical CFD Scores (with errors) Shifted Positive Following Independent Review



REFERENCES

1. Morris JC. Clinical dementia rating: a reliable and valid diagnostic and staging measure for dementia of the Alzheimer type. Int Psychogeriatr. 1997;9 Suppl 1:173-6; discussion 177-8. PubMed PMID: 9447441
2. Tractenberg RE, Schafer K, Morris JC. (2001). Interobserver disagreements on clinical dementia rating assessment: interpretation and implications for training. Alzheimer Dis Assoc Disord. 2001 Jul-Sep;15(3):155-61. PMID: 11522933

DISCUSSION

- Because the CDR is utilized as a sole primary endpoint in trials of early symptomatic AD, and is a critical outcome measure in trials at other stages of disease, including prevention trials, and its scoring can be challenging, particularly in mild disease, there is a need to identify CDR administrations within clinical trials that have an increased probability of scoring errors to maintain data integrity and improve signal detection.
- This study demonstrated that it is feasible to identify "atypical" patterns of domain scoring within the milder range of disease severity that are indicative of elevated rates of scoring errors.
- Identification of "atypical" patterns provides the opportunity to trigger independent review algorithms based entirely upon the internal pattern of domain scores.
- Further work is needed to identify potential atypical patterns in more advanced disease, and to potentially exploring more sophisticated data analytic approaches to explore internal score patterns in this and other scales that reflect an increased probability of scoring errors.

MedAvante-ProPhase

A WIRB-Copernicus Group Company