

# Global Assessment Measures for Assessing Efficacy in Dementia Drug Trials

*Kenneth Rockwood*

**ABSTRACT:** The use of judgment-based global measures of clinical treatment effects has a long history in the regulatory approval of drugs for the treatment of dementia. This paper reviews current use of these measures, noting several practical difficulties with their implementation. Importantly, global measures have usually yielded general estimates of drug effects similar to psychometric test batteries, even though correlations between global and psychometric measures at the individual patient level is modest. Like psychometric tests, global measures can tell us that some clinically detectable effect is present but often yield only limited evidence about what those effects might be. Steps should be taken to improve the specificity of treatment effect description and to incorporate patient/caregiver preferences in global measures about disease treatment in dementia.

**RÉSUMÉ:** Mesures globales d'évaluation pour l'évaluation de l'efficacité dans les essais thérapeutiques sur la démence. D'un point de vue historique, des mesures globales des effets du traitement fondées sur le jugement ont été utilisées pour l'approbation de médicaments pour traiter la démence. Cet article revoit l'utilisation actuelle de ces mesures tout en signalant plusieurs difficultés pratiques associées à leur application. Il est important de noter que les mesures globales ont habituellement fourni des estimés généraux des effets thérapeutiques qui sont semblables aux batteries de tests psychométriques, même si la corrélation entre les mesures globales et psychométriques au niveau individuel est modeste. Comme les tests psychométriques, les mesures globales peuvent indiquer qu'il y a certains effets cliniques qui sont détectables, mais elles fournissent souvent des données limitées sur ce que pourraient être ces effets. La spécificité de la description des effets du traitement et l'intégration des préférences du patient/soignant devraient être améliorées dans les mesures globales du traitement de la démence.

Can. J. Neurol. Sci. 2007; 34: Suppl. 1 - S52-56

Global assessment measures have long been used in dementia drug trials. As proposed elsewhere,<sup>1</sup> global measures chiefly serve two related purposes: most importantly, they allow judgments to be made by experienced clinicians about whether treatment has been associated with clinically detected change at the individual patient level; in addition, they allow this judgment to be quantified.

## BACKGROUND: TYPES OF GLOBAL MEASURES AND HISTORY OF THEIR USE

Two types of global, judgment-based measures have been described: specified and unspecified.<sup>2</sup> Specified measures comprise descriptions of patient characteristics across the stages of untreated dementia, chiefly Alzheimer's disease. Thus, for example, the Global Deterioration Scale describes seven stages, from a person with no subjective complaint and no objective deficit, to a patient dying of dementia.<sup>3</sup> Other instruments, such

as the Clinical Dementia Rating (CDR) and the Functional Rating Scale provide stage-specific descriptions by symptom domain (e.g., memory, judgment, behaviour, function).<sup>4,5</sup> In this way, specified clinical assessment measures, following the format of a semi-structured interview, provide clinicians with a means of recognizing disease profiles. In the early days of drug trials for dementia, such measures were used both as efficacy outcome assessments and eligibility criteria as a means of specifying the stage of dementia.

From the Division of Geriatric Medicine, Dalhousie University, Halifax, NS, Canada.

RECEIVED OCTOBER 7, 2005. ACCEPTED IN FINAL FORM APRIL 22, 2006.

Reprint requests to: Division of Geriatric Medicine, Dalhousie University, 1421-5955 Veterans' Memorial Lane, Halifax, Nova Scotia, B3H 1C6, Canada.

In contrast to specified global measures, “unspecified” measures have the form of brief guidelines for the completion of a Likert scale, weighted at 4 (“no change”) and ranging from 1 (“very much better”) to 7 (“very much worse”). (Levels 2 and 3 and 5 and 6 allow for “much” and “a little” improvement and worsening, respectively.) More recently, somewhat more structured but still unspecified global measures have been proposed and can be classified as two types: efforts to increase the standardization of the Likert-based summary clinical judgements, such as the Alzheimer’s Disease Cooperative Study Global Clinical Impression of Change;<sup>6</sup> and individualized measures such as Goal Attainment Scaling (GAS).<sup>7</sup>

Global measures have had an essential role in the evaluation of new compounds. In 1990, the United States Food and Drug Administration guidelines required a global outcome measure to be one of two primary outcomes (the other was a psychometric test) in registration trials as the “ultimate test of the clinical utility of a drug’s antidementia effects”.<sup>8</sup> Over the years, just as the Alzheimer’s Disease Assessment Scale-Cognitive Subscale (ADAS-cog) has come to be the psychometric test, the global outcome is usually the Clinician’s Interview-Based Impression of Change with Caregiver input (CIBIC plus) or a like measure.<sup>9</sup>

The 1995 summary of Canadian guidelines for the development of antidementia therapies noted this background but raised the concern that the unspecified measures, in particular, had been imposed without empirical testing of their measurement properties.<sup>10</sup> It suggested that further empirical testing was needed if global measures were to continue to have such an important role in antidementia drug testing.

#### EXPERIENCE WITH UNSPECIFIED GLOBAL MEASURES

Despite the Canadian guidelines’ call for a better empirical understanding of the properties of global outcome measures, progress has been only modest. A 2000 review concluded that there is little published information about the reliability and validity of such scales, and the situation has not changed much since.<sup>11</sup> An important and challenging exception, however, is a study in which a videotaped version of the CIBIC was evaluated by raters.<sup>12</sup> For study purposes, the raters were deceived, being told that they were evaluating interviews taken at baseline and at 6 to 12 months later. In actuality, half of the interviews were shown to them in reverse order, which allowed the ratings on “true order” interviews to be compared with “reverse order” ratings. While absolute agreement was poor ( $\kappa=0.18$ ), reasonable agreement was observed for those patients who were worse, or the same ( $\kappa=0.51$ ). In other words, clinicians were better at agreeing when patients had deteriorated than when they had improved. In short, clinicians seem to view success as the reversal of the pattern of disease progression without treatment, even though the face validity of such a metaphor is questionable.<sup>13,14</sup>

As discussed elsewhere,<sup>1,15</sup> there is a dilemma when unspecified global measures give results at odds with other measures, as clinical judgment will to some extent depend on clinicians’ experience with a class of compounds. In this context, it is worth considering that the agent linopirdine, which was the first drug tested by the group that became the Consortium of Canadian Centres for Clinical Cognitive Research, showed no

effect on the global clinical measure, even though the effect size of the ADAS-cog was large enough to have been detectable.<sup>16,17</sup> Similarly, an early report with the *Ginkgo biloba* extract EGb 761 showed a difference in the ADAS-cog at a level large enough to have been clinically detectable, but here too the global clinical measure detected no difference.<sup>18</sup> Although *Ginkgo* has been the focus of additional study with respect to preventing disease expression,<sup>19</sup> the use of a global measure appears to have detected true inefficacy with both compounds,<sup>20</sup> at a time when the very modest (but clinically detectable) effects detected by the ADAS-cog required interpretation and verification. Interestingly, clinicians do indeed appear to use the CIBIC-plus to give context to psychometric testing. Qualitative studies of the notes that clinicians used in the CIBIC-plus interviews suggest that they tend to give priority to caregiver reports of function and behaviour over their own estimates of cognition.<sup>21,22</sup> This is especially the case when caregivers report that function has improved or that behaviour has declined. Although clinicians’ narratives potentially contain important insights into the response to enhanced cholinergic neurotransmission, this resource has been little exploited.<sup>15</sup>

As noted, Goal Attainment Scaling (GAS) can be considered as another judgment-based unspecified measure. Despite limited use to date, it offers lessons for the current debate.<sup>23,24</sup> At baseline, patients and family members are asked which of the problems they have with dementia they would most like to see resolved. For each problem the current state is described and a treatment goal is set. A range of expectations is scaled so that other levels (from “very much worse [than the present state]” to “very much better [than the present state]”) are also described. At subsequent interviews, the description that best matches the patient’s present state is scored. A standardized formula is used to produce a summary score that reflects the overall degree of goal attainment. Clearly, GAS is a technique for both measurement and management and also reflects the sort of clinical reasoning that informs daily practice. It has been a sensitive measure of detecting clinically important change and slightly more responsive to change than the ADAS-cog.<sup>24</sup>

#### EXPERIENCE WITH SPECIFIED GLOBAL MEASURES

The most commonly used and best studied specified global measure in dementia drug trials is the CDR.<sup>4,25-27</sup> Although originally intended as a staging tool, the CDR can be scored to give a global measure, suitable for the calculation of change scores. The so-called “sum-of-boxes” scoring measure allows the scores to range from 0-18.

The CDR has had important variations and adaptations, including the extension of its domains in the Functional Rating Scale,<sup>28</sup> and its adaptation for use in long-term care.<sup>29</sup> Progression to severe dementia on the CDR has also been used as an outcome in a so-called ‘survival’ design.<sup>30</sup> Perhaps the other most widely used specified global scale is the Gottfries-Brane-Steen scale.<sup>31,32</sup> It consists of subscales that measure intellectual (12 items), emotional (3 items), and items of self-care and other ADLs (6 items), as well as six items of behavioural and psychological symptoms of dementia, and has been used in many (chiefly European) studies of both cholinergic and non-cholinergic compounds, as reviewed elsewhere.<sup>1</sup>

## LESSONS FROM THE USE OF GLOBAL MEASURES IN DEMENTIA DRUG TRIALS

The experience, to date, with global clinical measures suggests: the measures identify treatment effects of about the same order of magnitude as the ADAS-cog;<sup>33-36</sup> they do not identify the same patients as benefiting;<sup>24</sup> and scepticism about their use is based as much on a belief that they should not be as good as more objective measures, as it is about the empirical experience.

The latter concern has prompted a move to, if not eliminate global clinical measures then,<sup>3</sup> at least, standardize the interview on which the global impression of change is made.<sup>37,38</sup> Thus, for example, in the Alzheimer's Disease Cooperative Study Clinician's Global Impression of Change (ADCS CGIC) 15 domains were specified for assessment in the clinical interview, with change in each domain rated on a 7-point scale.<sup>6</sup>

While some standardization of the clinical interview might be helpful, in practice it means that an important opportunity to understand patient preferences (which appears to have been present with the unstructured CIBIC-plus) is lost. Moreover, as argued elsewhere, the move to standardization stems from an attempt to impose a particular measurement ethos on a subject matter (expert clinical judgment) to which it may well be unsuited.<sup>13,14,39</sup> All global measures share an essential feature: they incorporate clinical judgments. It might therefore be argued that they are particularly suited to studying new—and decidedly clinical—phenomena, such as treated Alzheimer's disease.<sup>40</sup> Judgment-based global measures share several essential features, including that the items selected for inquiry, and their weightings, should use patients' reports of what is troubling them, rather than coming from a theoretical scheme. Items should be easy to use and score, and heterogeneous, so that they capture all symptoms that are important to the phenomenon being investigated (rather than giving priority to homogeneity); similarly, all items should be included, rather than statistical criteria being used to eliminate items. Such measures can be contrasted with "psychometric" measures in which the opportunity for judgment is minimized. Despite having strong proponents,<sup>41</sup> judgment-based global measures also have their detractors who argue that advanced psychometric techniques have now caught up with whatever insights might have been afforded by such analyses.<sup>42</sup> Whether these advanced techniques have informed dementia drug studies seems doubtful, however.

Some of the scepticism about global clinical measures is rooted in a misapprehension about the nature of so-called "objective" tests in the evaluation of dementia. If there is one area of inquiry that should be sceptical about objective tests, it is dementia. For years, we have accepted neuropathology as the "gold standard" for objective evaluation, only to realize, now, that many people with "objectively" demonstrated neuropathological dementia are not, in fact, clinically demented.<sup>43</sup> Indeed, more recent attempts to make sense of the order in which plaques and tangles occur, and the order in which these lesions make sense, have relied on summary global clinical impressions to validate the hypothesis about the "objective" findings.<sup>44</sup> In short, for the foreseeable future, psychometric tests and even biomarkers will rely as much on validation against clinical measures as will the latter need to show some relationship with the former.

Accepting that while unspecified measures have problems with reliability, and that some of the standardization that specified global measures achieve comes at important costs, what should be recommended for the future use of such scales? It seems critical that we should not eliminate clinical judgment from understanding disease treatment effects. Employing – and systematically recording – clinical judgments allows us to understand whether treatment meets the preferences of patients and caregivers, so it also seems reasonable to use measures that do not leave these opportunities to chance.<sup>15</sup> In consequence, a measure such as GAS, which offers the potential to understand patient preferences when used in conjunction with an instrument like the ADCS-GGIC, would provide a non-arbitrary way to know whether a given trial produced clinically meaningful results. That these viewpoints do not always coincide is well illustrated in studies of GAS in which clinicians were more sceptical about claims of success than were patients/caregivers.<sup>23,24</sup> Global measures also challenge us to formulate a model of treatment success. Note that while clinical raters are sceptical about seeing improvement in patients treated with cholinesterase inhibitors, even when that improvement conforms to a "reversal of progression" model,<sup>12</sup> a recent meta-analysis suggests that the impression of benefit conferred with a global measure is similar, in the aggregate, to that conferred by the ADAS-cog.<sup>35</sup> In short, whatever individual patient assessment difficulties might exist, group effects are detectable. The problem – and this holds for the ADAS-cog – is that whatever is being detected is not readily translated into clinical practice.<sup>45</sup> In consequence, not only does a strict reliance on standardized tests leave us with the pragmatic problem of what to tell physicians to look for as treatment effects in practice, but we also are missing opportunities to gain the insights into cholinergic neurotransmission that would come from aggregating careful descriptions of individual results. These concerns are not theoretical. A survey of experienced physicians suggests that much of what they evaluate in clinical practice is not captured by the measures now employed in drug trials, save the CIBIC-plus, and, even there, it is captured in a way that is not readily aggregated.<sup>42</sup>

Capturing important information from the responses of patients to trials also offers an opportunity to use the trials as a means of better understanding brain function. For example, the cholinesterase inhibitor studies offer the possibility for new insights into cholinergic neurobiology. As the lists of agents that are employed to combat the various types of dementia grows, so too will our opportunities to learn about the neurobiology of such quintessentially human attributes as being recognizably one's self.

In addition, future trials can better be used to test our instrumentation. Any trial that employs more than one measure can be viewed as one in which patients and treatments are held constant but the measures are varied. From this, we can empirically address questions of responsiveness to change. In this context, an item of some interest will be to understand what it is about patients who are seen not to benefit on the ADAS-cog but who are, nevertheless, rated as having improved globally. Better research using global measures offers us the ability to make better judgments about who benefits from treatment.

## ACKNOWLEDGEMENTS

Kenneth Rockwood receives career support from the Canadian Institutes of Health Research through an Investigator award, and from the Dalhousie Medical Research Foundation as the Kathryn Allen Weldon Professor of Alzheimer Research.

## DECLARATION

Kenneth Rockwood receives research support from the CIHR and research funding from the Alzheimer Society of Canada. In the last five years, he has received speaker or consulting honoraria from Eisai, Janssen-Ortho, Lundbeck, Merck, Novartis and Pfizer, and research support from Janssen-Ortho, Merck and Pfizer.

## REFERENCES

- Rockwood K. Global assessment measures in dementia drug trials. In: Rockwood K, Gauthier S, editors. *Trial designs and outcomes in dementia therapeutic research*. London: Taylor & Francis; 2006: p.75-84.
- Rockwood K, Morris JC. Global staging measures. In: Gauthier S, editor. *Alzheimer's disease: diagnosis and treatment*. London: Martin Dunitz; 1996: p. 141-54.
- Reisberg B, Ferris SH, de Leon MJ, Crook T. The global deterioration scale for assessment of primary degenerative dementia. *Am J Psychiatry*. 1982;139:1136-9.
- Morris JC. The clinical dementia rating (CDR): current version and scoring rules. *Neurology*. 1993;43:2412-4.
- Feldman H, Schulzer M, Wang S, Tuokko H, Beattie BL. The functional rating scale in Alzheimer's disease assessment: a longitudinal study. In: Iqbal K, Mortimer JA, Winblad B, Wisniewski HM, editors. *Research advances in Alzheimer's disease and related disorders*. Chichester, UK: Wiley; 1995: p. 235-41.
- Schneider LS, Olin JT, Doody RS, Clark CM, Morris JC, Reisberg B, et al. Validity and reliability of the Alzheimer's disease cooperative study-clinical global impression of change. *The Alzheimer's disease cooperative study*. *Alzheimer Dis Assoc Disord*. 1997;11(Suppl 2):S22-32.
- Kiresuk TJ, Smith RE, Cardillo JE. *Goal attainment scaling: applications theory and measurement*. Hillside, NJ; Lawrence Erlbaum Associates; 1994.
- Leber PD. Developing safe and effective anti-dementia drugs. In: Becker R, Giacobini E, editors. *Alzheimer disease: from molecular biology to therapy*. Boston: Birkhauser; 1997.
- Rosen WG, Mohs RC, Davis KL. A new rating scale for Alzheimer's disease. *Am J Psychiatry*. 1984;41:1356-64.
- Mohr E, Feldman H, Gauthier S. Canadian guidelines for the development of antidementia therapies: a conceptual summary. *Can J Neurol Sci*. 1995;22:62-71.
- Oremus M, Perrault A, Demers L, Wolfson C. Review of outcome measurement instruments in Alzheimer's disease drug trials: psychometric properties of global scales. *J Geriatr Psychiatry Neurol*. 2000;13:197-205.
- Quinn J, Moore M, Benson DF, Clark CM, Doody R, Jagust W, et al. A videotaped CIBIC for dementia patients: validity and reliability in a simulated clinical trial. *Neurology*. 2002;58:433-7.
- Rockwood K. Use of global assessment measures in dementia drug trials. *J Clin Epidemiol*. 1994;47:101-3.
- Rockwood K, Wallack M, Tallis R. Treating dementia: understanding success short of cure. *Lancet Neurol*. 2003;2: 630-3.
- Rockwood K, Joffres C. Improving clinical descriptions to understand the effects of dementia treatment: consensus recommendations. *Int J Geriatr Psychiatry*. 2002;17:1006-11.
- Rockwood K, Beattie BL, Eastwood MR, Feldman H, Mohr E, Pryse-Phillips W, et al. A randomized, controlled trial of linopirdine in the treatment of Alzheimer's disease. *Can J Neurol Sci*. 1997;24:140-5.
- Rockwood K, Stolee P. Responsiveness of outcome measures used in an antidementia drug trial. *Alzheimer Dis Assoc Disord*. 2000;14:182-5.
- Le Bars PL, Katz MM, Berman N, Itil TM, Freedman AM, Schatzberg AF. A placebo-controlled, double-blind, randomized trial of an extract of Ginkgo biloba for dementia. *North American EGB Study Group*. *JAMA*. 1997;278:1327-32.
- van Dongen M, van Rossum E, Kessels A, Sielhorst H, Knipschild P. Ginkgo for elderly people with dementia and age-associated memory impairment: a randomized clinical trial. *J Clin Epidemiol*. 2003;56:367-76.
- Kurz A, Van Baelen B. Ginkgo biloba compared with cholinesterase inhibitors in the treatment of dementia: a review based on meta-analyses by the Cochrane collaboration. *Dement Geriatr Cogn Disord*. 2004;18:217-26.
- Joffres C, Graham J, Rockwood K. A qualitative analysis of the clinician interview-based impression of change (Plus): methodological issues and implications for clinical research. *Int Psychogeriatr*. 2000;12:403-13.
- Joffres C, Bucks RS, Haworth J, Wilcock GK, Rockwood K. Patterns of clinically detectable treatment effects with galantamine: a qualitative analysis. *Dement Geriatr Cogn Disord*. 2003;15:26-33.
- Rockwood K, Graham J, Fay S. Goal setting and attainment in Alzheimer's disease patients treated with donepezil. *J Neurol Neurosurg Psychiatry*. 2002;73:500-7.
- Rockwood K, Fay S, Song X, MacKnight C, Gorman M. Video-imaging synthesis of treating Alzheimer's disease (VISTA) investigators. Attainment of treatment goals by people with Alzheimer's disease receiving galantamine: a randomized controlled trial. *CMAJ*. 2006;174:1099-105.
- Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL. A new clinical scale for the staging of dementia. *Br J Psychiatry*. 1982;140:566-72.
- Morris JC, Ernesto C, Schafer K, Coats M, Leon S, Sano M, et al. Clinical dementia rating training and reliability in multicentre studies: the Alzheimer's Disease Cooperative Study experience. *Neurology*. 1997;48(6):1508-10.
- Rockwood K, Strang D, MacKnight C, Downer R, Morris JC. Inter-rater reliability of the clinical dementia rating in a multicentre trial. *J Am Geriatr Soc*. 2000;48:558-9.
- Feldman H, Gauthier S, Hecker J, Vellas B, Subbiah P, Whalen E. Donepezil MSAD Study Investigators Group. A 24-week, randomized, double-blind study of donepezil in moderate to severe Alzheimer's disease. *Neurology*. 2001;57:613-20.
- Marin DB, Flynn S, Mare M, Lantz M, Hsu MA, Laurans M, et al. Reliability and validity of a chronic care facility adaptation of the clinical dementia rating scale. *Int J Geriatr Psychiatry*. 2001;16:745-50.
- Sano M, Ernesto C, Thomas RG, Klauber MR, Schafer K, Grundman M, et al. A controlled trial of selegiline, alpha-tocopherol, or both as treatment for Alzheimer's disease. *The Alzheimer's Disease Cooperative Study*. *N Engl J Med*. 1997;336:1216-22.
- Gottfries CG, Brane G, Gullberg B, Steen G. A new rating scale for dementia syndromes. *Arch Gerontol Geriatr*. 1982;1:311-30.
- Brane G, Gottfries CG, Winblad B. The Gottfries-Brane-Steen scale: validity, reliability, and application in anti-dementia drug trials. *Dement Geriatr Cogn Disord*. 2001;12(1):1-14.
- Livingston G, Katona C. How useful are cholinesterase inhibitors in the treatment of Alzheimer's disease? A number needed to treat analysis. *Int J Geriatr Psychiatry*. 2000;15:203-7.
- Livingston G, Katona C. The place of memantine in the treatment of Alzheimer's disease: a number needed to treat analysis. *Int J Geriatr Psychiatry*. 2004;19:919-23.
- Whitehead A, Perdomo C, Pratt RD, Birks J, Wilcock GK, Evans JG. Donepezil for the symptomatic treatment of patients with mild to moderate Alzheimer's disease: a meta-analysis of individual patient data from randomised controlled trials. *Int J Geriatr Psychiatry*. 2004;19:624-33.

36. Rockwood K. Size of the treatment effect on cognition of cholinesterase inhibition in Alzheimer's disease. *J Neurol Neurosurg Psychiatry*. 2004;75:677-85.
37. Schneider LS, Ölin JT. Clinical global impressions in Alzheimer's clinical trials. *Int Psychogeriatr*. 1996;8:277-88.
38. Knopman DS. Global change assessments in anti-Alzheimer clinical drug trials. *Dement Geriatr Cogn Disord*. 1998; 9(Suppl 3): S8-15.
39. Rockwood K, MacKnight C. Assessing the clinical importance of statistically significant improvement in anti-dementia drug trials. *Neuroepidemiology*. 2001;20:51-6.
40. Feinstein AR. *Clinimetrics*. New Haven, CT: Yale University Press; 1987.
41. De Vet HC, Terwee CB, Bouter LM. Current challenges in clinimetrics. *J Clin Epidemiol*. 2003;56:1137-41.
42. Streiner D. Clinimetrics vs. psychometrics: an unnecessary distinction. *J Clin Epidemiol*. 2003;56:1142-5.
43. Polvikoski T, Sulkava R, Myllykangas L, Notkola IL, Niinisto L, Verkkoniemi A, et al. Prevalence of Alzheimer's disease in very elderly people: a prospective neuropathological study. *Neurology*. 2001;56:1690-6.
44. Bennett DA, Schneider JA, Wilson RS, Bienias JL, Arnold SE. Neurofibrillary tangles mediate the association of amyloid load with clinical Alzheimer's disease and level of cognitive function. *Arch Neurol*. 2004;61:378-84.
45. Rockwood K, Black S, Robilliard A, Lussier I. Potential treatment effects of donepezil not detected in Alzheimer's disease clinical trials: a physician survey. *Int J Geriatric Psychiatry*. 2004;19: 954-60.