International Psychogeriatrics, Vol. 12, No. 3, 2000, pp. 403-413

Qualitative Analysis of the Clinician Interview-Based Impression of Change (Plus): Methodological Issues and Implications for Clinical Research

CHRISTINE JOFFRES, JANICE GRAHAM, AND KENNETH ROCKWOOD

ABSTRACT. The Clinician Interview-Based Impression of Change, plus carer interview (CIBIC-Plus), is widely used in antidementia drug trials. It comprises Likert scales for disease severity and changes, and written accounts summarizing semistructured interviews evaluating behavior, cognition, and function. Studies using the CIBIC-Plus have focused on the numeric scores to the exclusion of the textual data. Our study explored both sets of data to evaluate whether the CIBIC-Plus written data supported (a) the clinicians' global evaluation of patients' changes during treatment, and (b) the emergence of consistent treatment effects. The global (numeric) scales of change were inconsistently supported by the textual data provided in the CIBIC-Plus. No consistent treatment effects were noted. Methodological problems presently limit the retrospective use of the CIBIC-Plus textual data. Improved standardization of note-taking in the CIBIC-Plus textual data may allow for a better understanding of the typical profiles and clinical importance of changes seen in the course of dementia treatment.

In 1990, the Division of Neuropharmacological Drug Products of the U.S. Food and Drug Administration recommended that antidementia drug trials include clinical global measures of change as primary efficacy outcome measures. These multidimensional measures are based on the premises that (a) a clinically useful drug must have a clinical effect, not only a cognitive one, and (b) the clinical effectiveness of new treatments should be apparent to experienced clinicians (Leber, 1997; Reisberg et al., 1997; Rockwood, 1994; Schneider & Olin, 1996).

A widely used version of the Clinician Interview-Based Impression of Change (CIBIC-Plus) includes textual data on the patients' history, general appearance, mental cognitive state, behavior, functional ability, and 7-point Likert scales recording disease severity and changes during and/or at the end of treatment

From the Queen Elizabeth II Health Sciences Centre, Halifax, Nova Scotia, Canada (C. Joffres; J. Graham; and K. Rockwood, MD).

Offprints. Requests for offprints should be directed to Kenneth Rockwood, MD, Queen Elizabeth II Health Sciences Centre, Geriatric Medicine Research Unit, Room 1421, 5955 Veterans' Memorial Lane, Halifax, Nova Scotia B3H 2E1, Canada. e-mail: rockwood@is.dal.ca

Received November 4, 1999
Accepted March 3, 2000

(Olin et al., 1996). These scales have been shown to have face validity (Knapp et al., 1994; Schneider et al., 1997) and predictive validity (Schneider et al., 1997). Knapp and colleagues (1994) and Schneider and coworkers (1997) demonstrated that change scales were sensitive to longitudinal change in 24- and 30-week studies. As well, Schneider and colleagues (1997) found that, at 12 months, the change scores of patients' global scales (N = 306) were significantly associated with the change scores of the Clinical Dementia Rating, Global Deterioration Scale, Mini-Mental State Examination, and Functional Assessment Staging (CDR, GDS, MMSE, and FAST). Similarly, in two different studies, Morris and colleagues (1998) and Cummings and coworkers (1998) found that metrifonate-treated patients exhibited significantly better scores on both the Alzheimer's Disease Assessment Scale-Cognitive (ADAS-Cog) and the CIBIC-Plus (a global change scale) than the placebo group, thereby further confirming the predictive validity of clinical change scales. This profile has held in several other studies in dementia (Bodick et al., 1997; Burns et al., 1999; Corey-Bloom et al., 1998; Rogers et al., 1998a, 1998b).

The criteria underlying global (numeric) scores of patients' changes have nevertheless had little formal evaluation, and studies that have used global measures of change have focused on the numeric scores to the exclusion of the written data. The goals of this study were to determine how CIBIC-Plus written data related to the clinicians' global evaluation of patients' changes, and whether identifiable treatment effects were present. In addition, this article presents some of the methodological issues that we encountered during the analyses of the textual data.

METHODS

Patients

Patient data (N = 18) came from a 6-month, Phase III, double-blind, randomized, placebo-controlled, multicenter trial of metrifonate in patients with probable Alzheimer's disease (AD) of borderline to marked severity. At baseline, 1 patient was borderline, 3 were mildly ill, 8 moderately ill, 4 markedly ill, and 2 had no reported global assessment. Patients' files were selected purposefully (Glaser & Strauss, 1967; Strauss & Corbin, 1990). The selection criteria for the files were relevancy (i.e., the information collected from the participants had to be relevant to our study) and richness of information (i.e., data had to be as specific and detailed as possible). Thus, we selected those files (18 of 42 available Canadian files from a 6-month multicenter trial) that included as much information as possible about the patients' history, cognition, behavior-mood, function, and changes during treatment.

Data Collection

Our data included the written texts and change scores of the patients' CIBIC-Plus files at baseline (BL), and at Visits 2, 3, and 4 (respectively called V2, V3, and V4). CIBIC-Plus forms are divided into five domains: the patient's recent clinical and social history (Domain 1), clinicians' observations about the patient's general appearance (Domain 2), the patient's mental cognitive state (Domain 3), behavior (Domain 4), and activities of daily living (ADLs) (Domain 5). The last three domains include several categories. Mental cognitive state has six: arousal-alertness-attention-concentration (a single category),

orientation, memory, language-speech, praxis, and judgment-problem solving-insight. Behavior includes: thought content, hallucinations-delusions-illusions, behavior-mood, sleep-appetite, and neurological-psychomotor activity. ADLs contain two categories: basic and complex functional ability, and social function. Each category includes probes guiding data collection and blank space for note-taking. The CIBIC-Plus forms were completed by clinicians experienced in treating AD, during semistructured interviews with patients and their caregivers at BL, 12 weeks after BL (V2), 18 weeks after BL (V3), and 26 weeks after BL (V4). A 2-week tolerance within this schedule was permitted.

Change scores were rated on a 7-point Likert scale. Rating of changes was as follows: 0 = not assessed, 1 = very much improved, 2 = much improved, 3 = minimally improved, 4 = no change, 5 = minimally worse, 6 = much worse, 7 = very much worse. Change scales were filled out at V2, V3, and V4. Rating of changes was to be made in reference to the BL visit.

Analyses

Textual data were analyzed qualitatively. Data were transcribed and imported into QSR NUD*IST (Qualitative Solutions and Research Non-numerical Unstructured Data Indexing Searching and Theory-building software) and coded, i.e., broken down into meaningful pieces and assigned a code. Codes or categories were partly defined by the CIBIC-Plus domains (i.e., relevant history, observation-evaluation, mental-cognitive state, behavior, and ADLs). Subnodes representing the different areas of each domain or subdomain (e.g., memory, praxis) were attached to these categories. Additional subnodes that illuminated the data in ways not provided by the already existing nodes were created as needed. Transcripts were reviewed several times to ensure that all relevant data were accounted for and systematically coded under the appropriate categories.

Data were analyzed through different processes that included pattern identification, clustering of conceptual groupings, identification of relationships between variables (Miles & Huberman, 1994a, 1994b), and constant comparisons (Strauss & Corbin, 1990, 1994). Patients' changes during treatment, their characteristics, and the contexts of changes were systematically identified, coded, and compared. Anomalous cases were contrasted with the rest of the data. Similarities and contrasts between categories and domains were explored both intraindividually and interindividually. Improvements and declines were displayed in tables along with their characteristics, the time frame at which they were noted (when the data were available), and their impact on other categories and/or domains. Finally, changes derived from the textual analyses were compared with their corresponding change scores.

To increase the reliability of our findings, we presented our data, coding procedures, categories, and initial results to a panel of international experts, including qualitative researchers and geriatricians, during a 2-day meeting. Additionally, the primary analyst regularly consulted with other qualitative researchers and local geriatricians throughout the data analyses.

RESULTS

No consistent treatment effects were identified in the cases under review, even

though we selected the most relevant and rich files. Instead, we observed inconsistencies in the recording of the textual information, and perhaps more importantly, discrepancies in how clinicians conceptualized expected levels of change under treatment. In addition to problems of legibility, several aspects of the format, content, and recording of the interviews appear to underlie the following inconsistencies:

File Format

The format of the patients' files was not consistent. Although we had a total of 42 files, we had to limit our analyses to 18 because 24 files did not contain textual notes after BL. As well, only 11 of 18 files included both the CIBIC-Plus forms and the CIBIC-Plus scales. The seven remaining files contained the CIBIC-Plus scales and some notes on patients' changes on separate sheets. The paucity of notes on patients' changes after BL compromised BL-subsequent visit comparisons.

File Content

There was marked variability in the length and content of the files. The most complete files (n = 11) were 35-40 pages long. They included information on several aspects of cognition, behavior, functional ability, and social activities. The seven remaining files, the length of which varied from seven to nine pages, contained only a few notes on a limited number of areas of cognition (three or four areas versus six in the longer files) and one aspect of patients' behavior (generally mood) versus five in the longer files. Shorter files rarely included information on concentration, sleep, appetite, psychomotor activity, ADLs, hobbies, or social activities.

Sources of Information. There were inconsistencies in the sources of information, which included patients, carers, and clinicians in the long files. However, the order in which carers and patients were interviewed was often unclear. The short files rarely noted carers' data. Instead, they primarily recorded cognitive tests.

Tests. As expected from an individualized scale, different clinicians used different tests to assess the patients' cognition and/or tested different aspects of the patients' cognition. For example, in 11 cases, clinicians systematically tested the patients' fluency and word generation whereas others did not. Similarly, in eight cases, clinicians tested patients' visuospatial, ideomotor, and ideational abilities, whereas others only assessed the patients' visuospatial abilities. Even when clinicians tested the same abilities, they sometimes collected data in different ways. As well, tests to evaluate the patients' ability to solve problems ranged from proverb clarifications, to Wechsler Adult Intelligence Scale (WAIS) comprehension problems, emergency judgment scenarios, clarification questions, and/or different combinations of any of the above tests. Physicians also used different scoring methods. Finally, it appears that, even when tested, patients were not tested at each visit.

Follow-Up on Patients' Changes. The qualitative analyses of the data revealed other issues that limited findings. First, the lack of textual data after BL made comparative analyses of the textual data and change scales at V2, V3, and V4 difficult in many cases. In addition, clinicians did not necessarily follow up on the symptoms that they noted at BL. For example, of the 11 clinicians who explored patients' ability to perform instrumental ADLs at BL, only 7 followed up on these activities during

treatment. In addition, clinicians tended to include less information on patients' changes at each visit.

Specificity. Lack of specificity was evident in several areas and impeded assessments of change. Terminology that is too general (e.g., "irritable at times," "spends much time searching," "STM declined," "repetitions," "patient is becoming more dependent all the time," "cannot fix broken furniture") cannot lend itself to specific interpretation. In addition, the onset of new and/or already existing symptoms was often unclear. Similarly, it was also often difficult to know if textual assessments of further declines and/or improvements at the end of treatment were based on comparisons with BL or V3 (i.e., the patient's previous visit). It appeared that many comparisons were not based on BL but on the patients' previous visit, in spite of the CIBIC-Plus written instructions.

Discrepant Information. Conflicting information between interviewer testing and informant reports was noted (e.g., "patient's praxis unchanged but patient spills and breaks more things," "patient's ADLs unchanged but patient participates more in household chores, e.g., make beds, etc."). Importantly, the conflicting data between clinicians and caregivers (e.g., a caregiver indicated that a patient's stuttering had improved whereas the clinician recorded that it had become worse), patients and caregivers, or caregivers over time (e.g., the same caregiver may give conflicting information about a patient's previous symptoms at two different visits) were often not systematically explored.

Clinicians' Assessments of Changes

The clinicians' understandings of change scores appeared to vary among physicians. For example, study clinicians (Cl, C2, and C3) seemed to have different definitions of "minimal improvement" (Table 1). For C1, one improvement in one aspect of one symptom (e.g., short-term memory [STM] or mood) seemed sufficient for a patient to be assessed as "minimally improved," whether or not the patient's symptomatology included declines. In addition, improvements only needed to come from one source of improvement. This contrasts with C2 and

TABLE 1. Clinicians' Definitions of Minimal Improvement at V2

C1	• *Only one* clinically meaningful improvement or a better score than at BL on a neuropsychological test. • Deteriorations in other areas can be present and do not appear to affect the minimal improvement assessment. • Number of deteriorations can be higher than the number of improvements. • *Only one* source of information necessary.
C2 & C3[a]	• *At least two* clinically meaningful improvements. • Deteriorations in other areas can be present and do not appear to affect the minimal improvement assessment as long as the number of deteriorations is equal to or inferior to the number of improvements and/or stable symptoms. • *Two congruent* sources of information with respect to improvements and declines.

Note. V2 = Visit 2; BL = baseline.
[a]C2 and C3 worked at the same site.

C3, for whom minimal improvement meant that the patients had improved in more than one area and that the number of declines was inferior to the number of improvements and/or stable symptoms. As well, for these physicians, information about recorded improvements or declines had to come from two congruent sources of information. Table 1 summarizes the above.

Discrepancies were also noted in the clinicians' definitions of "no change." C1 appeared to define no change in two ways. First, patients with no cognitive decline (as demonstrated through neuropsychological tests) during and/or at the end of treatment with or without declines in other areas were assessed as stable. Second, patients who had "perplexing on and off demonstrations" with respect to their cognitive abilities were also classified in the "no change" category. For C2, no change seemed to mean that the number of clinically meaningful stable symptoms was higher than the number of declines and/or improvements, that the patient had stable symptoms in the cognitive, behavioral, and functional domains, and that the stable symptoms came from two congruent sources of information. C3's definition of no change was similar to that of C2, except that patients did not need to have stable symptoms in the functional domain. Table 2 summarizes the clinicians' understanding of no change.

The clinicians' definitions of deterioration were also inconsistent, as illustrated in Table 3. Interestingly, C2's definition of stability seemed identical to his and C3's second definition of "minimally worse." As well, minimal improvement assessments did not require as many indications of things having changed as did assessments of minimal declines. This was particularly true for C1, whose patients were assessed as minimally improved, even though some of the patients' files included more declines (at least numerically) than improvements.

As well, in some records, patients' symptoms at V3 and V4 were not compared with their symptoms at BL, but with their symptoms at the previous visit. Practically, this meant that, for two of the study clinicians, patients assessed as stable ("no change") after a

TABLE 2. Clinicians' Definitions of No Change at V2

C1	• Patients' cognitive abilities have remained unchanged when assessed via tests. • Deteriorations in other areas can be present and do not appear to affect the clinician's assessment. • Only one source of information necessary. • **Or** Patients' results on cognitive ability tests on a specific day show a wide variability.
C2	• The number of clinically meaningful stable symptoms at V2 is higher than the number of declines and/or improvements. • Patient's stable symptoms include the cognitive, behavioral, and functional domains. • There are two congruent sources of information with respect to the stable symptoms.
C3	• The number of clinically meaningful stable symptoms is higher than the number of declines and/or improvements. • Patient's stable symptoms do not necessarily include the functional domain. • There are two congruent sources of information with respect to the stable symptoms.

Note. V2 = Visit 2.

TABLE 3. Clinicians' Definitions of "Minimally Worse" at V2

C1	• One decline only is necessary for a minimally worse assessment. However, patients with both declines and improvements were assessed as minimally improved, even if they had experienced only one improvement and more declines than improvements. • Recorded declines are limited to the cognitive and/or behavioral domains. • Only one source of information necessary.
C2 & C3	• Patient exhibits more declines than stable symptoms and/or improvements. • **Or** Patient exhibits an equal number of or less declines than stable symptoms. • In both cases, declines include the cognitive, behavioral, and functional domain. • There are two congruent sources of information with respect to the declines and stable symptoms.

Note. V2 = Visit 2.

prior evaluation of minimal improvement or decline had respectively improved or declined when compared to BL and would have been assessed as "minimally improved" or "minimally worse" by the third clinician, who systematically compared the patients' symptoms with their symptoms at BL. Similarly, because two of the study clinicians compared declines at each visit to the patients' state at the prior visit, successive scores of minimally worse actually reflected considerable deterioration from BL. Such discrepancies were not witnessed in the case of improvement, where the referent seemed more consistently to be the BL.

The clinicians' lack of consistency with respect to their bases of comparisons when making assessments of patients after V2 raises some questions. Whereas two of the clinicians' assessments of no change at V3 after an assessment of minimally improved or minimally worse at V2 meant that the patients were better or worse at V3 than at BL, the same assessment by the third clinician meant that the patients' AD had remained stable since BL. Further complicating the issue, C2 and C3's assessments of minimally improved after a previous assessment of minimally worse did not necessarily mean that the patient was better at V3 than at BL. Similarly, C2 and C3's assessments of minimally worse at V3 after a previous assessment of minimally improved did not necessarily mean that the patients' AD was worse when compared with BL.

DISCUSSION

Qualitative analyses of the CIBIC-Plus files revealed inconsistencies in the format, content, referents, data collection, and, apparently, the clinicians' model of what represents a treatment effect. In general, improved assessments required fewer signs of improvements than declined assessments required signs of deteriorations, which might be due to the fact that improvements tend to be less expected in the course of AD than deteriorations.

The reliability of the data is weakened not only by the textual data lack of specificity but also by problems of within-file inconsistencies. The retrospective design of the study makes it difficult to

know the extent to which these inconsistencies reflect problems in the CIBIC-Plus and its implementation or problems in the note-taking. Inasmuch as these files came from a study in which the CIBIC-Plus global scores correlated with the ADAS-Cog in demonstrating a treatment effect, there is reason to suspect that poor note-taking may explain some of our findings. On the other hand, poor interrater reliability may reflect the clinicians' discrepancies around the definitions of change scores and criteria of comparisons to evaluate changes. We believe it likely that these discrepancies reflect discrepant understandings of exactly what to expect in patients with partially treated AD. This should not be unexpected. The total experience of any physician presently is more with untreated than with treated AD. In contrast to well-defined staging systems for untreated AD (Morris, 1997; Reisberg et al., 1982), there is no systematic model of AD treatment.

This study is exploratory, and attempts at generalization may be premature at this point. However, our discussions with the expert panel, and other investigators in Canada, the United States, and the United Kingdom lead us to suspect that the data recorded in CIBIC notes widely lack specificity and consistency (e.g., symptoms were not systematically recorded at follow-up), that the clinicians' understanding of rating scores differs from clinician to clinician, and that the boundaries between rating scores often overlap. Partly this would seem to be due to clinicians widely using the notes as an *aide-mémoire* rather than as data records, and partly this likely reflects true variability in the approach to dementia and its treatment.

Further attempts at generalization should be based on additional studies. It is important to remember that the above definitions have been established from the data that we have and that more detailed information about the patients' symptoms might have changed our understanding of the clinicians' definitions of no change, minimally improved, and minimally worse.

Our point is not to discredit clinicians' assessments of patients' AD. Quite the contrary, we believe that the patients' change scores reflected the patients' evolving symptomatology, as was borne out by the CIBIC-Plus change scores in the metrifonate studies (Cummings et al., 1998; Morris et al., 1998). It is important that consistency of the textual information in the patients' files allows individual patient problems to be tracked over the course of treatment if we are to better understand treatment effects. More detailed and specific texts may have allowed us to determine whether metrifonate had consistent effects with respect to specific domains (e.g., behavior, cognition) and/or specific areas of these domains (e.g., mood).

It is also important to caution that written clinical assessments need to capture those aspects of the clinical interview process that best ensure validity: individualization and contextualization of the data. For example, the interpretation of a standard test showing impaired visuospatial function in a retired ship's navigator who was also a talented painter will have to be different from the interpretation of the same result of impaired visuospatial function in a bookkeeper of no known artistic bent. The point of AD treatment is not to produce "standard" cognitive function but to be

of help to individuals in overcoming their particular deficits. Similarly, as has been pointed out elsewhere (Rockwood, 1994; Schneider & Olin, 1997), a great merit of the CIBIC-Plus is that it does not impose an a priori, and necessarily untested, version of what constitutes a treatment effect, but rather appropriately allows that to be defined on a case-by-case basis. Although a price necessarily is paid in apparent reliability, this is preferable to a highly reliable, structured questionnaire of uncertain validity in a given case. An unspecified model of AD treatment requires an unspecified measure. As noted above, however, internal consistency of observations is preferable and not at odds with the need for an individualized/contextualized approach.

The following recommendations would further strengthen the validity of studies based on global measures of change:

File Format

The file format should be consistent across participating physicians. Clinicians should use similar forms and have a clear understanding of how and when these forms have to be filled, while at the same time individualizing and giving context to changes in patients' AD.

File Content

Clinicians should be aware of the type and amount of information to be included in the files. Operational definitions of the CIBIC-Plus probes may facilitate the clinicians' understanding of the probes and provide a greater consistency with respect to the recorded data. Consistent follow-up on the patients' symptoms should be emphasized. There is no merit in following an exhaustive symptom inventory at every visit in the name of standardization. Rather, these symptoms that are important to individual patients should be recorded at BL (or as they emerge) and subsequently tracked.

Clinicians' *initial* descriptions of patients' deficits need to be specific and include information about the frequency, duration, extent/scope/severity, and intensity of symptoms at BL. Similarly, because changes can take many forms, the dimensional properties of changes must be clearly recorded. Again, these should include information about the frequency, duration, extent/scope/severity, and intensity of symptoms. A new category titled "New symptoms" or "Other changes" may allow clinicians to record changes that may not fit into any of the already existing categories.

Consistency with respect to the sources of information (who and what are the sources of information? When and for what to use them?) and the order in which to use them is also important. Previous research has demonstrated that the order in which informants (patients and carers) were interviewed tended to influence the clinicians' ratings of change scales (Reisberg et al., 1995). Test scoring methods should be uniform across participating clinicians.

Change Score Descriptors

CIBIC-Plus global change scores need better qualitative descriptors of change. As noted by Knopman and colleagues (1994), the current lack of descriptors "probably impairs its [the CIBIC] reliability" (p. 2320). As well, the current lack of clear criteria with respect to the change scores (e.g., minimally improved, very much improved, minimally worse,

etc.) makes the CIBIC-Plus score changes hardly more meaningful to researchers or other interested clinicians than a 3- or 4-point change on the MMSE.

As more effective treatments of AD become available, there is a need to go beyond the specialist physicians who have participated in drug development if treatments are to be more widespread. The CIBIC-Plus texts have the potential to usefully inform treating physicians about treatment effects and to translate the study results into everyday practice. There is a pressing need for a better understanding of typical treatment effects so that descriptors can emerge.

REFERENCES

Bodick, N. C., Offen, W. W., Levey, A. I., Cutler, N. R., Gauthier, S. G., et al. (1997). Effects of xanomeline, a selective muscarinic receptor agonist, on cognitive function and behavioral symptoms in Alzheimer's disease. *Archives of Neurology, 54,* 465-473.

Burns, A., Rossor, M., Hecker, J., Gauthier, S., Petit, H., et al. (1999). The effects of donepezil in Alzheimer's disease—Results from a multinational trial. *Dementia and Geriatric Cognitive Disorders, 10,* 237-244.

Corey-Bloom, J., Anand, R., & Veach, J. (1998). A randomized trial evaluating the efficacy and safety of ENA 713 (rivastigmine tartrate), a new acetylcholinesterase inhibitor, in patients with mild to moderate severe Alzheimer's disease. *International Journal of Geriatric Psychopharmacology, 1,* 55-65.

Cummings, J. L., Cyrus, P. A., Bieber, F., Mas, J., Orazem, J., et al. (1998). Metrifonate treatment of the cognitive deficits of Alzheimer's disease. *Neurology, 50,* 1214-1221.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research.* Chicago: Aldine.

Knapp, M. J., Knopman, D. S., & Solomon, P. R. (1994). A 30-week randomized controlled trial of high-dose tacrine in patients with Alzheimer's disease. *Journal of the American Medical Association, 271,* 985-991.

Knopman, D. S., Knapp, M. J., Gracon, S. L., & Davis, C. S. (1994). The Clinician Interview-Based Impression (CIBI): A clinician's global change rating scale in Alzheimer's disease. *Neurology, 44,* 2315-2321.

Leber, P. D. (1997). Developing safe and effective antidementia drugs. In R. Becker & E. Giacobini (Eds.), *Alzheimer disease: From molecular biology to therapy* (pp. 579-584). Boston: Birkhauser.

Miles, M. B., & Huberman, A. M. (1994a). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Newbury Park, CA: Sage.

Miles, M. B., & Huberman, A. M. (1994b). Data management and analysis methods. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 428-444). Thousand Oaks, CA: Sage.

Morris, J. C., Cyrus, P. A., Orazem, J., Mas, J., Bieber, F., et al. (1998). Metrifonate benefits cognitive, behavioral, and global function in patients with Alzheimer's disease. *Neurology, 50,* 1222-1230.

Morris, J. C., Ernesto, C., Schafer, K., Coats, M., Leon, S., et al. (1997). Clinical Dementia Rating training and reliability in multicenter studies: The Alzheimer's Disease Cooperative Study experience. *Neurology, 48,* 1508-1510.

Olin, J. T., Schneider, L. S., Doody, R. S., Clark, C. M., Ferris, S. H., et al. (1996). Clinical evaluation of global change in Alzheimer's disease: Identifying consensus. *Journal of Geriatric Psychiatry and Neurology, 9,* 176-180.

Reisberg, B., Schneider, L., Doody, R., Anand, R., Feldman, H., et al. (1997). Clinical global measures of dementia: Position paper from the international working group on harmonization of dementia drug guidelines. *Alzheimer Disease and Associated Disorders, 11*(3), 8-18.

Reisberg, B., Ferris, S. H., de Leon, M. J., & Crook, T. (1982). The Global Deterioration Scale for assessment of primary degenerative dementia. *American Journal of Psychiatry, 139,* 1136-1139.

Rockwood, K. (1994). Use of global assessment measures in dementia drug trials. *Journal of Clinical Epidemiology, 47,* 101-103.

Rogers, S. L., Doody, R. S., Mohs, R. C., & Friedhoff, L. T. (1998a). Donepezil improves cognition and global function in Alzheimer disease: A 15-week, double-blind, placebo-controlled study. *Archives of Internal Medicine, 158,* 1021-1031.

Rogers, S. L., Farlow, M. R., Doody, R. S., Mohs, R., & Friedhoff, L. T. (1998b). A 24-week, double-blind, placebo-controlled trial of donepezil in patients with Alzheimer's disease. *Neurology, 50,* 136-145.

Schneider, L. S., & Olin, J. T. (1996). Clinical global impression of change: Clinical global impression of change in Alzheimer's clinical trials. *International Psychogeriatrics, 8,* 277-290.

Schneider, L. S., Olin, J. T., Doody, R. S., Clark, C. M., Morris, J. C., et al. (1997). Validity and reliability of Alzheimer's disease cooperative study-clinical global impression of change. *Alzheimer Disease and Associated Disorders, 11*(Suppl. 2), S22-S32.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques.* Newbury Park, CA: Sage Publications.

Strauss, A., & Corbin, J. (1994). Grounded theory methodology: An overview. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 273-285). Thousand Oaks, CA: Sage Publications.

Acknowledgments. The authors are grateful to Bayer Canada for supporting this study, and to Raewyn Bassett for her comments on an earlier version of the manuscript.