

## Research Article

# FACE-TO-FACE VERSUS REMOTE ADMINISTRATION OF THE MONTGOMERY-ASBERG DEPRESSION RATING SCALE USING VIDEOCONFERENCE AND TELEPHONE

Kenneth A. Kobak, Ph.D.,<sup>1\*</sup> Janet B.W. Williams, D.S.W.,<sup>1,2</sup> Elizabeth Jeglic, Ph.D.,<sup>1,3</sup> Donna Salvucci, M.Ed.,<sup>1</sup> and Ian R. Sharp, Ph.D.<sup>1</sup>

*Although the use of telemedicine in psychiatry has a long history in providing clinical care to patients, its use in clinical trials research has not yet been commonly employed. Telemedicine allows for the remote assessment of study patients, which could be done by a centralized, highly calibrated, and impartial cohort of raters independent of the study site. This study examined the comparability of remote administration of the Montgomery-Asberg Depression Rating Scale (MADRS) by videoconference and by telephone to traditional face-to-face administration. Two parallel studies were conducted: one compared face-to-face with videoconference administration (N = 35), and the other compared face-to-face with telephone administration (N = 35). In each study, depressed patients were interviewed independently twice: once in the traditional face-to-face manner, and the second time by either videoconference or teleconference. A counterbalanced order was used. The mean MADRS score for interviews conducted remotely by videoconference was not significantly different from the mean MADRS scores conducted by face-to-face administration (mean difference = 0.51 points),  $P = .388$ , intraclass correlation (ICC) = .94,  $P < .0001$ . Similarly, the mean MADRS score for interviews conducted by telephone was not significantly different from the mean MADRS score conducted by face-to-face administration (mean difference = 0.74 points),  $P = .270$ , ICC = .93,  $P < .0001$ . Results of the study support the comparability of remote administration of the MADRS, by both telephone and videoconference, to face-to-face administration. Comparability of the administration mode allows for remote assessment of patients in both research and clinical applications. Depression and Anxiety 0:1–7, 2007. © 2007 Wiley-Liss, Inc.*

**Key words:** assessment; outcomes; telemedicine; depression; clinical trials; reliability and validity

## INTRODUCTION

Telemedicine is rapidly becoming an accepted mode of providing clinical services to patients [Mair and Whitten, 2000; Whitten and Love, 2005]. Eighteen states now pay for medical services provided remotely through telemedicine technologies under their state Medicaid programs [Johnson, 2006], following the lead of countries like Australia, where telemedicine has become an integral part of the health services delivery system [Hawker et al., 1998; Yellowlees and Kennedy, 1997], and is supported by national legislation [Kennedy

<sup>1</sup>MedAvante Research Institute, MedAvante Inc., Madison, Wisconsin

<sup>2</sup>Department of Psychiatry, Columbia University, New York, New York

<sup>3</sup>John Jay College of Criminal Justice, New York, New York

Contract grant sponsor: MedAvante, Inc.

\*Correspondence to: Kenneth A. Kobak, Ph.D., MedAvante Inc., 7601 Ganser Way, Madison, WI 53719.  
E-mail: kkobak@medavante.net

Received for publication 14 November 2006; Revised 24 July 2007; Accepted 27 July 2007

DOI 10.1002/da.20392

Published online in Wiley InterScience (www.interscience.wiley.com).

and Yellowlees, 2000]. Telepsychiatry (i.e., the use of telemedicine in psychiatry) has a long history of providing clinical care to patients [Baer et al., 1997], with patient satisfaction being generally positive [Hyler et al., 2005]. More recently, studies have found that remote clinical assessment of patients using rating scales such as the Hamilton Depression Scale (HAMD), Hamilton Anxiety Scale (HAM-A), Positive and Negative Syndrome Scale (PANSS), and Yale-Brown Obsessive Compulsive Scale (YBOCS) can be accomplished successfully using telemedicine, yielding similar results as when the scale is administered in the traditional face-to-face manner [Baer et al., 1995; Hyler et al., 2005; Kobak, 2004; Simon et al., 1993; Yoshino et al., 2001; Zarate et al., 1997].

These findings provide support for the use of telepsychiatry in clinical trials research, an application that has not yet been commonly employed. There are several ways in which telepsychiatry can improve clinical trials' methodology. Rater training and calibration can be delivered to diverse sites using videoconference technology [Kobak et al., 2006b, 2005b], allowing for direct observation of trainee's clinical interviews in real time, thus improving inter-rater reliability, a vexing problem in multicenter studies that has been cited as a possible cause for the increasing rate of failed trials [Muller and Szegedi, 2002; Perkins et al., 2000]. Remote administration by telephone enables the assessment of patients by clinicians at more frequent intervals, by obviating the need for travel to the study site. Another application is the use of centralized raters [Kobak et al., 2006a]. With this methodology, clinical trial patients can be assessed remotely by a centralized group of expert raters, who are highly calibrated to one another, and who can be blinded to study design and even study visit, reducing expectation bias. Centralized ratings also reduce the sheer number of raters involved, as all the study ratings can be conducted by a single rater cohort, enabling tighter calibration. Having the ratings divorced from the study site also minimizes the potential bias that might occur at baseline due to enrollment pressures, and the potential unblinding of raters due to collecting adverse event data. Centralizing raters also allows for better ongoing monitoring of interview quality, a dimension that was recently found to be associated with better signal detection [Kobak et al., 2005a].

This study examines the comparability of remote administration of the Montgomery-Asberg Depression Rating Scale [MADRS; Montgomery and Asberg, 1979] by videoconference and by telephone to traditional face-to-face administration. The MADRS is used widely as the primary outcome measure in clinical depression trials, both in the United States and globally. The MADRS is often chosen instead of the HAMD in clinical trials, as it is shorter, overcomes some of the methodological criticisms of the HAMD [e.g., poor item reliabilities and lack of internal construct validity; Bagby et al., 2004; Gibbons et al.,

1993], and shows comparable sensitivity to change over time [Khan et al., 2002].

Two parallel studies were conducted: one compared face-to-face with videoconference administration, and the other compared face-to-face with telephone administration. They are described in detail below.

## METHODS

### STUDY 1: VIDEOCONFERENCE VERSUS FACE-TO-FACE

Thirty-five subjects (12 men and 23 women, mean age = 44 years, range 21–66 years) with a DSM-IV-TR diagnosed mood disorder (major depression = 19; major depression in partial remission = 9; minor depression = 3; dysthymia = 1; bipolar disorder, depressed = 1; depression NOS = 2) were included. Diagnoses were determined using a modified version of the mood module of the Mini Neuropsychiatric Interview [Sheehan et al., 1998], and the overview section from the Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Patient Edition [First et al., 2002]. A range of mood disorders were included in order to evaluate the comparability of video and face-to-face administration across a wide range of symptom severity. The sample was 77% Caucasian, 14% African American, 3% Hispanic, and 6% others. Subjects were recruited through newspaper advertisements. All subjects signed informed consent statements approved by Allendale IRB, and were paid \$50 (US) for their participation.

**Research design.** Two cohorts or "waves" were used. In the first wave, subjects were interviewed with the MADRS by two different interviewers using the traditional face-to-face method. This was done to establish a "baseline" of the inter-rater reliability between the raters using the traditional face-to-face method, against which the reliability using videoconference administration could be compared. Each interviewer was blind to the other rater's ratings. Conducting independent (versus joint) evaluations of the same patient mitigates the artificial inflation of reliability coefficients that occurs when one rater interviews the patient and the second rater rates by simply observing the first rater's interview. Conducting independent evaluations is thus a more rigorous test of inter-rater reliability.

In the second wave, subjects were also interviewed by two different interviewers. However, in this cohort, one of the interviews was done face-to-face, and the other was conducted remotely using videoconferencing. Again, each rater was blind to the other rater's results. The same subjects (and raters) were used in both "waves" to control for patient difficulty and rater clinical skill as confounding factors. The order of which "wave" the subject received first was counter-balanced, as was the order of administration of face-to-face versus videoconference administration. Within

each wave, the two interviews were administered on the same day to control for changes due to time. To minimize memory effects, a distracting task was conducted between interviews. The time between waves was 1–3 days. After completing both waves of the study, the subjects completed a patient satisfaction survey, to evaluate feasibility and patient acceptance of the remote technology.

The rater cohort consisted of two male and four female interviewers. Five had doctoral degrees (four in psychology and one in social work), and one a masters degree in counseling psychology. Before the study, the raters underwent reliability training on the scale, consisting of a didactic review of the scale and scoring conventions, followed by at least three practice interviews (two group and one individual) observed by a trainer. Raters also observed each others' training sessions to enhance learning. They had a range of earlier experience with the MADRS, from extensive (K.K., J.B.W.W.) to minimal. The raters were paired using numerous permutations of dyads, to maximize generalizability. All the raters had earlier experience of conducting interviews remotely via videoconferencing.

**Remote videoconferencing technology.** Videoconference interviews were conducted using two H.323 IP standards-based Polycom iPower Videoconferencing Systems (Polycom; Pleasanton, CA). The two devices were connected via T1 line over a virtual private network, run at a minimum of an industry standard bit rate of 384kps. The system is HIPAA compliant, and occurs via a secure encrypted connection. The remote rater had the ability to control all functions of both the rater-side and the patient-side cameras, including zoom-out, and up/down/left/right movements. Camera pre-sets afforded the rater instant, one button, camera relocation to a predetermined position, such as patient head-and-shoulder or patient full-body views. Camera zoom was internal, so no camera movement was visible to the subjects.

## STUDY 2: TELEPHONE VERSUS FACE-TO-FACE

Thirty-five subjects (12 men and 23 women, mean age = 43 years, range 20–72 years) with a DSM IV-TR diagnosed mood disorder (major depression = 21; major depression in partial remission = 7; bipolar disorder, depressed = 1; minor depression = 1; dysthymia = 3; depression NOS = 2) were included. Subjects were recruited and diagnosed using the methods described in study 1. The sample was 86% Caucasian, 10% African American, 2% Hispanic, and 2% Asian.

In this study, subjects were administered the MADRS both by telephone and face-to-face in a counter-balanced order. Interviews were conducted on the same day to control for changes over time. A different

cohort of subjects was used for the telephone study than the videoconferencing study. The same group of raters was used, again in various dyad pairings.

Raters in both studies used a semi-structured interview guide [Structured Interview Guide for the MADRS, or SIGMA; Williams and Kobak, 2006]. The SIGMA preserves the original MADRS scale items, but includes questions to guide the clinician in their evaluation. Studies have found that using a structured interview guide improves reliability [Moberg et al., 2001]. For telephone administration, apparent sadness was rated based on non-visual signs like tone of voice, rate of speech, and crying as well as direct probes from the SIGMA interview (e.g., “In the past week, do you think you have looked sad or depressed to other people; Did anyone say you looked sad or down; How about when you’ve looked in the mirror; Did you look gloomy or depressed?”). This technique has been used successfully in self-report versions of the MADRS [Mundt et al., 2006] administered by computer as well as in computerized self-report versions of the HAMD [Kobak et al., 1990] and HAMA [Kobak et al., 1993].

## RESULTS

### STUDY 1: FACE-TO-FACE VERSUS VIDEOCONFERENCE ADMINISTRATION

The mean MADRS scores for interviews conducted remotely by videoconference was not significantly different from the mean MADRS scores conducted by face-to-face administration (21.74 versus 21.23, respectively,  $SD = 10.2$  and  $10.0$ , respectively, mean difference = 0.51 points),  $t(34) = 0.875$ ,  $P = .388$ . The intraclass correlation (ICC) between MADRSs administered by videoconference and those administered face-to-face was  $r = .94$ ,  $P < .0001$ . The correlation (ICC) between the 35 pairs of face-to-face interviews was  $r = .90$ ,  $P < .0001$ . There was no statistically significant difference between the two correlations, Fisher's  $z = 1.027$ ,  $P = .3044$ . That is, the correlation between one face-to-face and one remote interview was not significantly different from two interviews both done face-to-face.

Internal consistency reliability (Cronbach  $\alpha$ ) was also examined for the two modes of administration. Similar levels of internal consistency were found for interviews administered via videoconference ( $r = .90$ ) and those done face-to-face ( $r = .89$ ).

An examination of the inter-rater reliability (ICC) on the item level is presented in Table 1. The ICCs between videoconference and face-to-face were good to excellent for all items. The ICCs between videoconference and face-to-face were similar to the ICCs between two face-to-face interviews for most items.

There were no technical problems encountered during the study, e.g., dropped calls, freeze frames.

## STUDY 2: FACE-TO-FACE VERSUS TELEPHONE ADMINISTRATION

The mean MADRS score for interviews conducted by telephone (21.49,  $SD = 10.10$ ) was not significantly different from the mean MADRS score conducted by face-to-face administration (22.23,  $SD = 10.49$ ; mean difference = 0.74 points),  $t(34) = 1.121$ ,  $P = .270$ . The correlation (ICC) between MADRSs administered by telephone and those given face-to-face was  $r = .93$ ,  $P < .0001$ . The ICC between telephone and face-to-face for individual MADRS items is presented in Table 1. All item correlations were above .60, and significant at the  $P = .001$  level.

Internal consistency reliability (coefficient  $\alpha$ ) was also examined for the two modes of administration. The internal consistency for interviews administered via telephone ( $r = .899$ ) was comparable to those done face-to-face ( $r = .896$ ).

**Interview length.** The mean interview length conducted by videoconference (26.4 min,  $SD = 10.61$ ) was not significantly different from the mean length of those conducted face-to-face (25.6 min,  $SD = 10.14$ ),  $t(34) = 0.513$ ,  $P = .611$ . Similarly, there was no difference in interview length between those conducted by telephone (24.0 min,  $SD = 7.48$ ) and those conducted face-to-face (24.5 min,  $SD = 7.50$ ),  $t(31) = -0.332$ ,  $P = .742$ .

**Subject satisfaction.** When asked "How comfortable were you being interviewed", the percentage endorsing either "fairly comfortable" or "very comfortable" was 68% for video, 88% for telephone, and 85% for face-to-face administration,  $\chi^2(2) = 5.631$ ,  $P = .059$ . When asked "How much did you like being interviewed?", the percentage endorsing "liked a little" or "liked a lot" was 71% for video, 91% for telephone, and 94% for face-to-face,  $\chi^2(2) = 9.884$ ,  $P = .007$ . When asked "How well was the interviewer able to evaluate your symptoms and feelings?", the percentage endorsing "well" or "very well" was 91% for video, and

100% for both telephone and face-to-face,  $\chi^2(2) = 9.301$ ,  $P = .009$ . Thirty-five percent of subjects said the video interfered with their ability to communicate with the interviewer, compared with 52% for the telephone,  $\chi^2(1) = 1.795$ ,  $P = .180$ . Finally, when those in the video study were asked which method they preferred, 12% preferred video, 33% preferred face-to-face, and 55% had no preference,  $\chi^2(2) = 13.364$ ,  $P = .001$ . For those in the telephone study, 19% preferred the telephone, 34% preferred face-to-face, and 47% had no preference,  $\chi^2(2) = 5.719$ ,  $P = .057$ . All subjects said they would be willing to be interviewed again by either method (phone or videoconference).

## DISCUSSION

Results of the study support the comparability of remote administration of the MADRS, by both telephone and videoconference, to face-to-face administration. There was a high correlation between forms of administration, the mean score differences were not significantly different, and the internal consistencies were comparable. The correlation between two face-to-face interviews was not significantly different from the correlation between one face-to-face and one video or telephone administered interview, suggesting no diminished reliability when using remote administration, when holding the raters constant. Raters were successfully able to assess apparent sadness remotely by a combination of additional probes and reliance on vocal quality (in the case of the telephone), and videoconference image (in the case of video administration). Given the rather small sample size, the power was .80 to detect a population effect size of .49 for a paired  $t$  test with a two-tailed  $\alpha$  level of .05. On the basis of the standard deviations of the difference between video and face-to-face in the study sample, this would correspond to a difference of about 1.7 points on the MADRS. We acknowledge the sample

**TABLE 1. Intraclass correlation between video and face-to-face and between two face-to-face interviews on individual MADRS items**

	Video versus face-to-face ( $N = 35$ )	Face-to-face versus face-to-face ( $N = 35$ )	Telephone versus face-to-face ( $N = 35$ )
1. Apparent sadness	.78	.84	.87
2. Reported sadness	.84	.82	.89
3. Inner tension	.73	.61	.78
4. Reduced sleep	.88	.91	.80
5. Reduced appetite	.87	.85	.87
6. Concentration difficulties	.79	.93	.77
7. Lassitude	.72	.82	.73
8. Inability to feel	.81	.77	.81
9. Pessimistic thoughts	.69	.75	.67
10. Suicidal thoughts	.87	.92	.90
Total	.94	.90	.93

All ICCs significant at  $P < .001$ .

ICC, intraclass correlation; MADRS, Montgomery-Asberg Depression Rating Scale.

size does not provide adequate statistical power to detect a smaller difference. Nevertheless, such a difference would not be clinically meaningful.

The 95% confidence interval of the mean difference in study 1 was  $-0.68$  to  $+1.71$ . A two-point difference between scores is within the acceptable range or “margin or error” as reported in the literature for scales with similar range and constructs [Hamilton, 1967]. Memory effects may have contributed to the high ICCs due to the short (approximately 15 min to half an hour) time between testings, although these are similar to the ICCs found in the original validation study [Montgomery and Asberg, 1979], and to those found by others [Mundt et al., 2006]. Using a shorter interval between testings controls for subject changes, and a similar interval has been used with no memory effects [Kobak et al., 1993, 1990].

The findings support the work of Hermens et al. [2006], who found a high correlation between telephone and face-to-face administered MADRS, and only a half point difference between forms. The Hermens study did not include a traditionally administered control group; thus, it cannot be determined if a similar (or higher) correlation would be found with two face-to-face administrations. This study provided such a comparison, thus enabling a determination of the comparability of ICCs found by remote and face-to-face methods. It should be noted that overall ICC is related directly to the amount and efficacy of pre-study rater training and calibration [Targum, 2005]. Again, the key in determining comparability of mode of administration is the comparability of the ICCs, obtained by each method, regardless of the overall level obtained.

The comparability of the psychometric properties of remote to traditional face-to-face administration lends support for the use of remote clinical assessment in clinical trials. Recently, a large NIMH funded study (Sequenced Treatment Alternatives to Relieve Depression, STARD) used telephone-administered HAMDS as the primary outcome measure [Rush et al., 2004, 2006], assessing 2,876 subjects in 41 sites by a remote group of centralized clinicians. A study examining the comparability of videoconference versus face-to-face assessment of depression in a multi-center placebo-controlled trial is currently underway. Preliminary results confirm the hypothesis that using blinded remote-centralized raters results in significantly different study populations [Kobak et al., 2006a] and lower placebo response [Engelhart et al., 2007]. The blinded centralized raters generally scored depression severity lower than site raters at screen and baseline, with ratings coalescing at endpoint. Establishing the comparability of video and remote HAMDS before the study was critical to rule out the possibility that these differences were due to the mode of administration [i.e., video or face-to-face; Kobak, 2004]. Differences in signal detection between face-to-face ratings performed at the site and remote centralized ratings

conducted via videoconference will be examined when the blind is broken.

Given that both telephone and videoconference administrations were comparable to face-to-face administration, a key question is, what is the relative advantage of one remote method of administration over the other? Cukor et al. [1998] report on growing evidence that most of the clinical information is provided by the audio communication, whereas the video information mainly provides a “social presence.” Although some of this may be attributable to the lower bandwidth availability at the time of the review, more recent data support this hypothesis. For example, Strauss [unpublished manuscript] found more accurate assessment of job seekers’ intelligence, conscientiousness, and extraversion when interviews were conducted by telephone than when conducted by videoconference. Similarly, we found that raters were more accurate in guessing which subjects were real patients and which were actors when the interviews were conducted by teleconference than when they were conducted by videoconference [Kobak et al., 2006b]. In some disorders such as schizophrenia, where the evaluation of negative symptoms would not be feasible by telephone, the role of bandwidth plays an important role, with higher bandwidth essential for the accurate assessment of symptoms [Yoshino et al., 2001]. We are currently examining the use of remote videoconference assessment of schizophrenia in a large multi-center trial. Given the hypothesis that video provides a social connection that telephone does not provide, one would predict that patients would prefer video over telephone. In this study, we were surprised to find that the percentage of patients who preferred the telephone compared with face-to-face was greater than the percentage who preferred video to face-to-face. To some extent this may be explained by the fact that these were subjects recruited via newspaper ads specifically for the purposes of evaluating the comparability of the methodology, and they had no therapeutic relationship with the clinicians. In a similar study where the same subjects were evaluated by both telephone and videoconference, more subjects preferred the videoconference [Kobak et al., 2006b], despite reporting technological problems (slower bandwidth was used in that trial). In both studies, all patients reported being willing to be interviewed again using both technologies. Hilty et al. [1999] found that primary care patients preferred face-to-face to remote mental health evaluations for both initial assessment and follow-up appointments, suggesting that patients may have some reservation about the use of telemedicine technology. Other studies in primary care have found greater acceptance of remote telemedicine, and that this acceptance increases over time [Williams et al., 2007].

Future studies on the use of remote administration of clinical interviews by telephone or videoconference in placebo-controlled clinical trials are needed to examine the relative impact of this methodology on effect size.

The hypothesized benefits of increased reliability, quality, and objectivity that remote centralized ratings could provide need to be examined in terms of actual improvement in signal detection relative to traditional face-to-face administration. If these hypotheses are confirmed, remote administration may provide a tool enabling significant improvement in clinical trials methodology.

**Acknowledgments.** The authors would like to acknowledge the contributions of Dr. Karen Rybchenko for helping administer the clinical interviews conducted in the study, Dr. Kia Crittenden and Kristy Harris, M.A. for patient screening, and Don Borque and Peggy Barclay for assistance with study coordination. The funding for this study was provided by MedAvante, Inc. Dr. Kobak is VP of Research at MedAvante, and Dr. Williams is VP of Clinical Development at MedAvante.

## REFERENCES

- Baer L, Cukor P, Jenike MA, Leahy L, O'Laughlen J, Coyle JT. 1995. Pilot studies of telemedicine for patients with obsessive-compulsive disorder. *Am J Psychiatry* 153:1383-1385.
- Baer L, Elford DR, Cukor P. 1997. Telepsychiatry at forty: what have we learned? *Harv Rev Psychiatry* 5:7-17.
- Bagby RM, Ryder AG, Schuller DR, Marshall MB. 2004. The Hamilton Depression Rating Scale: has the gold standard become a lead weight? *Am J Psychiatry* 161:2163-2177.
- Cukor P, Baer L, Willis BS, Leahy L, O'Laughlen J, Murphy M, Withers M, Martin E. 1998. Use of videophones and low-cost standard telephone lines to provide a social presence in telepsychiatry. *Telemed J* 4:313-321.
- Engelhart N, Kobak KA, Leuchter A, Williams JBW, Cook I, Leon AC, DeBrotta DJ. 2007. Site vs. centralized raters in a clinical depression trial: impact on patient selection and placebo response. Boca Raton, FL: National Institute of Mental Health, New Clinical Drug Evaluation Unit, 47th Annual Meeting.
- First MB, Spitzer RL, Gibbon M, Williams JBW. 2002. Structured clinical interview for DSM-IV-TR Axis I disorders, research version, Patient Edition. (SCID-I/P). New York: Biometrics Research, New York State Psychiatric Institute.
- Gibbons RD, Clark DC, Kupfer DJ. 1993. Exactly what does the Hamilton Depression Rating Scale measure. *J Psychiatr Res* 27: 259-273.
- Hamilton M. 1967. Development of a rating scale for primary depressive illness. *Br J Soc Clin Psychiatry* 6:278-296.
- Hawker F, Kavanagh S, Yellowlees P, Kalucy RS. 1998. Telepsychiatry in South Australia. *J Telemed Telecare* 4:187-194.
- Hermens ML, Ader HJ, van Hout HP, Terluin B, van Dyck R, de Haan M. 2006. Administering the MADRS by telephone or face-to-face: a validity study. *Ann Gen Psychiatry* 5:3.
- Hilty DM, Servis ME, Nesbitt TS, Hales RE. 1999. The use of telemedicine to provide consultation-liaison service to the primary care setting. *Psychiatr Ann* 29:421-427.
- Hyler SE, Gangure DP, Batchelder ST. 2005. Can telepsychiatry replace in-person psychiatric assessments? A review and meta-analysis of comparison studies. *CNS Spectr* 10:403-413.
- Johnson K. 2006. TV Screen, not couch, is required for this session. *New York: New York Times*; 1p.
- Kennedy C, Yellowlees P. 2000. Guidelines for using videoconferencing services in mental health services. *J Telemed Telecare* 6:352-353.
- Khan A, Khan SR, Shankles EB, Polissar NL. 2002. Relative sensitivity of the Montgomery-Asberg Depression Rating Scale, the Hamilton Depression Rating Scale and the Clinical Global Impressions Rating Scale in antidepressant clinical trials. *Int Clin Psychopharmacol* 17:281-285.
- Kobak KA. 2004. A comparison of face-to-face and videoconference administration of the Hamilton Depression Rating Scale. *J Telemed Telecare* 10:231-235.
- Kobak KA, DeBrotta DJ, Engelhart N, Williams JBW. 2006a. Site vs. centralized raters in a clinical depression trial. Boca Raton, FL: National Institute of Mental Health, New Clinical Drug Evaluation Unit, 46th Annual Meeting.
- Kobak KA, Engelhardt N, Lipsitz JD. 2006b. Enriched rater training using internet based technologies: a comparison to traditional rater training in a multi-site depression trial. *J Psychiatr Res* 40:192-199.
- Kobak KA, Feiger AD, Lipsitz JD. 2005a. Interview quality and signal detection in clinical trials. *Am J Psychiatry* 162:628.
- Kobak KA, Reynolds WM, Greist JH. 1993. Development and validation of a computer-administered version of the Hamilton Anxiety Scale. *Psychol Assess* 5:487-492.
- Kobak KA, Reynolds WM, Rosenfeld R, Greist JH. 1990. Development and validation of a computer-administered version of the Hamilton Depression Rating Scale. *Psychol Assess* 2:56-63.
- Kobak KA, Williams JBW, Engelhardt N. 2005b. Face to face vs. remote assessment of inter-rater reliability on the HAMD. Boca Raton, FL: National Institute of Mental Health, New Clinical Drug Evaluation Unit, 45th Annual Meeting.
- Mair F, Whitten P. 2000. Systematic review of studies of patient satisfaction with telemedicine. *BMJ* 320:1517-1520.
- Moberg PJ, Lazarus LW, Mesholam RI, Bilker W, Chuy IL, Neyman I, Markvart V. 2001. Comparison of the standard and structured interview guide for the Hamilton Depression Rating Scale in depressed geriatric inpatients. *Am J Geriatr Psychiatry* 9:35-40.
- Montgomery SA, Asberg M. 1979. A new depression scale designed to be sensitive to change. *Br J Psychiatry* 132:382-389.
- Muller MJ, Szegedi A. 2002. Effects of interrater reliability of psychopathologic assessment on power and sample size calculations in clinical trials. *J Clin Psychopharmacol* 22:318-325.
- Mundt JC, Katzelnick DJ, Kennedy SH, Eisfeld BS, Bouffard BB, Greist JH. 2006. Validation of an IVRS version of the MADRS. *J Psychiatr Res* 40:243-246.
- Perkins DO, Wyatt RJ, Bartko JJ. 2000. Penny-wise and pound-foolish: the impact of measurement error on sample size requirements in clinical trials. *Biol Psychiatry* 47:762-766.
- Rush AJ, Fava M, Wisniewski SR, Lavori PW, Trivedi MH, Sackeim HA, Thase ME, Nierenberg AA, Quitkin FM, Kashner TM, Kupfer DJ, Rosenbaum JF, Alpert J, Stewart J, McGrath PJ, Biggs MM, Shores-Wilson K, O'Neal BL, Lebowitz BD, Ritz AL, Niederehe G, for the STAR\*D Investigators Group. 2004. Sequenced treatment alternatives to relieve depression (STAR\*D): rationale and design. *Control Clin Trials* 25:119-142.
- Rush AJ, Trivedi MH, Wisniewski SR, Stewart JW, Nierenberg AA, Thase ME, Ritz L, Biggs MM, Warden D, Luther JF, Shores-Wilson K, Niederehe G, Fava M for the STAR\*D Study Team. 2006. Bupropion-SR, sertraline, or venlafaxine-XR after failure of SSRIs for depression. *N Engl J Med* 354:1231-1242.
- Sheehan DV, Lecrubier Y, Sheehan K, Harnett A, Amorim P, Janavs J, Weiller E, Baker R, Dunbar G. 1998. The mini international neuropsychiatric interview (MINI): the development and validation of a structured diagnostic psychiatric interview for DSM IV and ICD-10. *J Clin Psychiatry* 59:22-23.

- Simon GE, Revicki D, von Korff M. 1993. Telephone assessment of depression severity. *J Psychiatr Res* 27:247–252.
- Targum SD. 2005. Rater competency for mood disorders scales. Boca Raton, FL: National Institute of Mental Health, New Clinical Drug Evaluation Unit, 45th Annual Meeting.
- Whitten P, Love B. 2005. Patient and provider satisfaction with the use of telemedicine: overview and rationale for cautious enthusiasm. *J Postgrad Med* 51:294–300.
- Williams JBW, Ellis A, Middleton A, Kobak KA. 2007. A pilot study of video telemedicine in the assessment of primary care patients with anxiety and depressive disorders. Boca Raton, FL: National Institute of Mental Health, New Clinical Drug Evaluation Unit, 47th Annual Meeting.
- Williams JBW, Kobak KA. 2006. Structured interview guide for the Montgomery and Asberg Depression Rating Scale (SIGMA). New York: New York State Psychiatric Institute.
- Yellowlees P, Kennedy C. 1997. Telemedicine: here to stay. *Med J Aust* 166:262–265.
- Yoshino A, Shigemura J, Kobayashi Y, Nomura S, Shishikura K, Den R, Wakisaka H, Kamata S, Ashida H. 2001. Telepsychiatry: assessment of televideo psychiatric interview reliability with present- and next-generation internet infrastructures. *Acta Psychiatr Scand* 104:223–226.
- Zarate CA Jr, Weinstock L, Cukor P, Morabito C, Leahy L, Burns C, Baer L. 1997. Applicability of telemedicine for assessing patients with schizophrenia: acceptance and reliability. *J Clin Psychiatry* 58:22–25.