

▶ A comparison of face-to-face and videoconference administration of the Hamilton Depression Rating Scale

Kenneth A Kobak

Research Training Associates and MedAvante, Madison, Wisconsin, USA

Summary

To determine whether the mode of administration affected the psychometric properties of the Hamilton Depression Rating Scale (HAMD), 21 subjects with an affective disorder were administered two HAMD interviews on the same day, once via videoconference (at 384 kbit/s) and once face-to-face, by two independent interviewers. The interviewers were blind to the results of the other interview. The mean HAMD scores were almost identical (18.4 for videoconferencing and 18.2 for face to face). The intra-class correlation between the two sets of scores was 0.88. When another group of 21 subjects had the HAMD administered in two face-to-face interviews on the same day, the inter-rater reliability was not significantly different from that for the videoconference cohort. Most patients (91%) thought videoconferencing was a useful way to receive a psychological evaluation when other means were unavailable or limited. The study shows the psychometric equivalence of remote and face-to-face HAMD administration.

Introduction

Psychiatric services have been provided by videoconference for over a decade^{1,2}. Several studies have shown that videoconferencing is an effective means of linking patients in remote locations to psychiatrists^{3,4} and of facilitating communication between mental health providers^{5,6}. In Australia, videoconferencing has become an integral part of mental health services^{7,8} and is supported by national legislation⁴. The method has many applications, such as the remote provision of expert legal testimony for hearings concerning involuntary hospitalization⁹ and the linking of low-income inner-city patients to teaching hospitals¹⁰. Patient acceptance of the method has generally been good^{5,11}. In one study, even patients with schizophrenia (with delusions of reference involving the television) found the method was acceptable, without exacerbation of their delusions¹².

While great strides have been achieved in the use of videoconferencing in clinical practice, its use in clinical trials has been more limited. Outcome measures in trials of psychiatric treatments are typically scores on rating scales, obtained in clinical interviews with patients. Before videoconferencing can be used in clinical trials that make use of rating scales, it must be shown that the scales' psychometric properties are not affected by their remote administration.

Some evidence to support this was provided by Baigent *et al.*¹³, who found high levels of agreement between diagnoses obtained remotely and those obtained face to face, as well as high levels of inter-rater reliability for ratings of severity of illness made using the Brief Psychiatric Rating Scale. In another study, Ball and Puffett¹⁴ found high correlations between face-to-face and remote (videoconference) assessments of cognitive function in an elderly sample. Menon *et al.*¹⁵ found no significant difference between the total scores on the Hamilton Depression Rating Scale (HAMD) obtained by videoconference and by face-to-face administration, although the study was too small for significance testing ($n=12$) and item scores were not reported.

Accepted 29 April 2004

Correspondence: Dr Kenneth Kobak, Research Training Associates and MedAvante, 7601 Ganser Way, Madison, WI 53719, USA
(Fax: +1 608 829 1411; Email: Kobak@charter.net)

The present study was carried out to examine the equivalence of face-to-face and remote administration of the HAMD. The HAMD¹⁶ is the most widely used outcome measure in psychiatric clinical trials and has become the 'gold standard' against which other rating scales are validated¹⁷. Demonstrating that HAMD scores obtained via videoconference are equivalent to HAMD scores obtained in the traditional, face-to-face manner will support the remote administration of the HAMD in clinical trials.

Methods

Subjects

Forty-two subjects (22 men, 20 women; mean age 37 years, SD 12, range 18–62; 36 Caucasian, 4 African-American, 1 Hispanic and 1 Asian) were recruited through newspaper advertisements (patients in US clinical trials are typically recruited in this manner). Subjects were screened by telephone using the mood module of the Mini International Neuropsychiatric Interview (MINI)¹⁸ and the question on bipolar disorder from the PRIME-MD instrument¹⁹. Those who met the American Psychiatric Association's criteria²⁰ for an affective disorder were eligible to participate. Subjects had the following diagnoses: major depression ($n=28$); minor depression ($n=1$); depression in partial remission ($n=7$); bipolar disorder, depressed ($n=4$); depression not otherwise specified ($n=1$); and bipolar disorder not otherwise specified ($n=1$). All subjects with an affective disorder were included, in order to examine the comparability of face-to-face and videoconference HAMD administration across a wide spectrum of disease.

All subjects signed informed consent statements and were paid \$50 for their participation (\$1 is €0.8). Ethics approval was obtained from the appropriate committee.

Procedure

There were two cohorts of 21 subjects. The first cohort was interviewed with the HAMD twice by two different interviewers using the traditional, face-to-face method. This was done to establish the inter-rater reliability between the two raters using the face-to-face method. Each interviewer was blind to the results of the other rater's results.

The second cohort of 21 subjects was also interviewed twice by two different interviewers. 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

order of administration was balanced between interviewers.

The raters were one male and one female interviewer, the former with extensive experience with the HAMD and the latter relatively new to the technique. Both raters had graduate degrees in psychology.

In order to control for time effects, cohorts were balanced, so that the first patient was from cohort 1, the second from cohort 2, and so on. In both cohorts, both evaluations were conducted on the same day in order to control for changes in the severity of the affective disorder over time. In order to minimize memory effects, a distracting task was conducted between interviews. Finally, after completion of both interviews, subjects in the second cohort completed a patient satisfaction survey, to evaluate the feasibility and patient acceptance of videoconferencing.

Videoconferencing equipment

Standard commercial videoconferencing units were used (iPower Standard, Polycom). The units, on different floors of the same building, were directly connected at 384 kbit/s via an Ethernet cable. The remote rater could control both the local and the patient cameras.

Results

The mean HAMD scores obtained remotely by videoconference and face to face were almost identical (18.4 vs 18.2 respectively, mean difference=0.15 points; $P=0.817$) (Table 1). The inter-rater reliability (calculated as the intraclass correlation coefficient²¹) was 0.88 ($P<0.001$, 95% CI 0.74 to 0.95). Similarly, the inter-rater reliability between the 21 pairs of interviews where both were conducted face to face was 0.93 ($P<0.0001$, 95% CI 0.84 to 0.97). There was no significant difference between the two correlations ($z=-0.85$, $P=0.20$, 95% CI for the difference -0.73 to 0.36). That is, the inter-rater reliability between face-to-face and remote interviews was not significantly different from that obtained when the two interviews were done face to face. The differences between

Table 1 HAMD scores obtained in videoconferences and face to face

	Mean (SD)	<i>n</i>	Range
Videoconference	18.4 (5.3)	21	8–27
Face to face	18.2 (6.0)	21	8–29

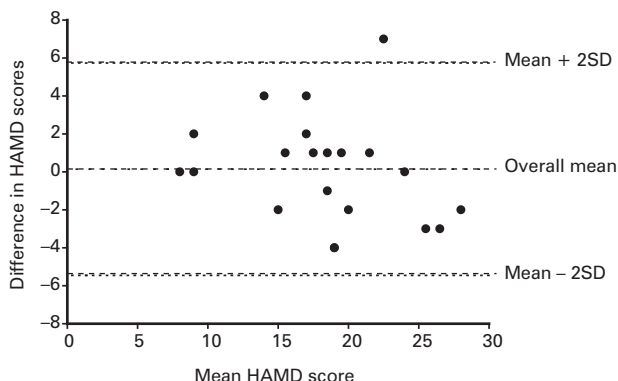


Fig 1 Differences between the HAMD scores for the two modalities and their means ($n=21$ — note that two of the data points overlap).

videoconference and face-to-face HAMD scores are compared against the mean difference scores in Fig 1.

An examination of the inter-rater reliability on the item level is presented in Table 2. As can be seen, the inter-rater reliability between videoconference and face-to-face modes of administration was high and significant for most items. Of particular interest is the correlation on the two observational items, psychomotor retardation and psychomotor agitation. The inter-rater reliability for retardation was reasonably high, at $r=0.59$ ($P=0.002$), while that for psychomotor agitation was low, at $r=0.20$ ($P=0.18$). The intra-class correlations between the two face-to-face interviews and those between videoconference and face-to-face interviews were similar for most items.

Internal consistency reliability (alpha coefficients) was also examined for the two modes of administration. Similar levels of internal consistency were found for both versions ($r=0.74$ for videoconferencing, $r=0.77$ for face-to-face administration).

Patient satisfaction

Following the testing session, subjects who received a videoconference HAMD interview were asked to evaluate their experience with the technology. Of the 21 subjects, 14 (67%) reported feeling fairly or very comfortable being interviewed remotely, 15 (71%) liked the experience 'a little or a lot' and 19 (91%) thought this was a useful way to receive a psychological evaluation when other means were unavailable or limited. Only 7 (33%) reported that the technology interfered with their ability to communicate with the interviewer. Sixteen (76%) said they would be willing to be interviewed via videoconference in order to avoid travelling for an office visit, and 15 (71%) said they would like to be interviewed via videoconference again.

Discussion

Before a rating scale can be used routinely in a mode other than the one originally intended, it must be shown that this will not affect the scores obtained, or its psychometric properties. This is of particular concern in research applications, where comparability of results between studies that have used the scale is

Table 2 Intraclass correlation coefficients for individual HAMD items

	Videoconference vs face to face ($n=21$)	Face to face vs face to face ($n=21$)	All subjects combined ($n=42$)
1. Depressed mood	0.77***	0.80***	0.78***
2. Guilt	0.68***	0.07	0.32*
3. Suicide	0.87***	0.99***	0.94***
4. Initial insomnia	0.82***	0.88***	0.85***
5. Middle insomnia	0.88***	0.94***	0.90***
6. Terminal insomnia	0.48**	0.63***	0.54***
7. Work and interests	0.16	0.70***	0.48***
8. Psychomotor retardation	0.59**	0.68***	0.63***
9. Psychomotor agitation	0.20	0.40*	0.25*
10. Psychic anxiety	0.55**	0.61**	0.58***
11. Somatic anxiety	0.60**	0.90***	0.79***
12. Somatic, gastrointestinal	0.97***	0.73***	0.86***
13. Somatic, general	0.69***	0.58**	0.63***
14. Sexual interest	0.69***	0.97***	0.85***
15. Hypochondriasis	0.63***	0.51**	0.55***
16. Loss of weight	0.95***	0.93***	0.94***
17. Insight ^a	—	—	—
<i>Total score</i>	<i>0.88***</i>	<i>0.93***</i>	<i>0.91***</i>

^aAll patients scored zero on item 17.
* $P<0.05$, ** $P<0.01$, *** $P<0.001$.

critical. The results of the present study support the equivalence of face-to-face and videoconference administration of the HAMD. The mode of administration did not affect the inter-rater reliability nor the internal consistency of the scale, the correlation between the scores obtained in the two modes of administration was high, and the mean scores obtained by each method of administration were almost identical.

The results support the findings of Menon *et al.*¹⁵, who found no significant difference between the HAMD scores when the instrument was administered face to face or by video-phone, although that sample was small ($n=12$). Similarly, the present results support the work of Simon *et al.*²², who found that HAMD scores obtained by telephone were similar to those obtained face to face. The addition of video should not only increase accuracy—as Menon *et al.* wrote, ‘the added value of video creates a social presence that enables the conferencing partners to know each other better and feel more comfortable discussing complex issues’¹⁵.

The high correlation between videoconference and face-to-face HAMD scores on the psychomotor retardation item provides support for the ability of videoconferencing to capture non-verbal cues and accurately to show psychomotor retardation. In contrast, correlations for psychomotor agitation were less robust. One possible reason for this is that retardation can be inferred to a great extent from facial expression and rate of speech, while agitation is based primarily on motor movement. More careful attention to the full body picture during the videoconference may be helpful. It should also be noted that the low reliability of the agitation item has been consistently reported in the literature, even with face-to-face administration; for example, Williams²³ found a correlation of 0.11 when using the SIGHD structured interview guide, and correlations of 0.20 and 0.32 when using an unstructured interview.

In addition to psychometric equivalence, subjects’ acceptance of the method is also an important factor in determining the feasibility of a scale. Subjects may decline an invitation to participate in a clinical trial if they dislike being interviewed in the manner required. In our study, patients found the videoconferences generally acceptable and 76% said they would be willing to use videoconferencing in order to avoid travelling to the doctor’s office for a visit. The high level of acceptance was especially noteworthy as the subjects were recruited from the community and had no prior relationship with the interviewers. When conducted within the context of an existing professional relationship, acceptance would likely be even greater. As with all innovations, as the technology

improves and becomes more widely used, assessment via videoconferencing should become even more readily accepted.

The use of videoconferencing for psychiatric assessment in clinical trials has several possible applications. Patients could be rated remotely from their physician’s office. This would allow primary-care physicians to become partners in the research process; as primary-care physicians are the principal provider of mental health services, this would probably be an effective method of recruiting patients for a trial. It would also improve both the generalizability of the findings and the representativeness of patients participating in clinical trials, many of whom are at present atypical or treatment resistant and so have high drop-out rates and high placebo response rates²⁴. Another possible application is the use of centralized groups of expert raters, who could remotely conduct the assessments for all the subjects in a multi-site trial. This should improve signal detection by reducing the variability inherent with large, diverse groups of raters. As videoconferencing technology improves and becomes more widely available, its use in clinical research can be expected to increase. Determining that administration via videoconferencing does not affect the psychometric properties of assessments traditionally done face-to-face should facilitate the use of such scales in telemedicine work.

References

- 1 Yellowles P, McCoy WT. Telemedicine. A health care system to help Australians. *Medical Journal of Australia* 1993;**159**:437–8
- 2 Preston J, Brown FW, Hartley B. Using telemedicine to improve health care in distant areas. *Hospital and Community Psychiatry* 1992;**43**:25–32
- 3 Kennedy C, Yellowles P. The effectiveness of telepsychiatry measured using the Health of the Nation Outcome Scale and the Mental Health Inventory. *Journal of Telemedicine and Telecare* 2003;**9**:12–16
- 4 Kennedy C, Yellowles P. Guidelines for using videoconferencing in mental health services. *Journal of Telemedicine and Telecare* 2000;**6**:352–3
- 5 McLaren P, Ball CJ, Summerfield AB, Watson JP, Lipsedge M. An evaluation of the use of interactive television in an acute psychiatric service. *Journal of Telemedicine and Telecare* 1995;**1**:79–85
- 6 Yellowles P, Kennedy C. Telemedicine applications in an integrated mental health service based at a teaching hospital. *Journal of Telemedicine and Telecare* 1996;**2**:205–9
- 7 Hawker F, Kavanagh S, Yellowles P, Kalucy RS. Telepsychiatry in South Australia. *Journal of Telemedicine and Telecare* 1998;**4**:187–94
- 8 Yellowles P, Kennedy C. Telemedicine: here to stay. *Medical Journal of Australia* 1997;**166**:262–5
- 9 Yellowles P. The use of telemedicine to perform psychiatric assessments under the Mental Health Act. *Journal of Telemedicine and Telecare* 1997;**3**:224–6
- 10 McLaren PM, Laws VJ, Ferreira AC, O’Flynn D, Lipsedge M, Watson JP. Telepsychiatry: outpatient psychiatry by videolink. *Journal of Telemedicine and Telecare* 1996;**2** (suppl. 1):59–62
- 11 Ball CJ, McLaren PM, Summerfield AB, Lipsedge MS, Watson JP. A comparison of communication modes in adult psychiatry. *Journal of Telemedicine and Telecare* 1995;**1**:22–6

- 12 Dongier M, Tempier R, Lalinec-Michaud M, Meunier D. Telepsychiatry: psychiatric consultation through two-way television. A controlled study. *Canadian Journal of Psychiatry* 1986;**31**:32–4
- 13 Baigent MF, Lloyd CJ, Kavanagh SJ, *et al.* Telepsychiatry: 'tele' yes, but what about the 'psychiatry'? *Journal of Telemedicine and Telecare* 1997;**3** (suppl. 1):3–5
- 14 Ball C, Puffett A. The assessment of cognitive function in the elderly using videoconferencing. *Journal of Telemedicine and Telecare* 1998;**4** (suppl. 1):36–8
- 15 Menon AS, Kondapavalru P, Krishna P, *et al.* Evaluation of a portable low cost videophone system in the assessment of depressive symptoms and cognitive function in elderly medically ill veterans. *Journal of Nervous and Mental Disease* 2001;**189**:399–401
- 16 Hamilton M. A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry* 1960;**23**:56–62
- 17 Kobak KA. The Hamilton Depression Inventory. In: Maruish ME, ed. *Handbook of Psychological Assessment in Primary Care Settings*. Mahwah, NJ: Lawrence Erlbaum Associates, 2000
- 18 Sheehan DV, Lecrubier Y, Sheehan KH, *et al.* The Mini International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM–IV and ICD–10. *Journal of Clinical Psychiatry* 1998;**59** (suppl. 20): 22–3
- 19 Spitzer RL, Williams JBW, Kroenke K, *et al.* Utility of a new procedure for diagnosing mental disorders in primary care. The PRIME-MD 1000 study. *Journal of the American Medical Association* 1994;**272**:1749–56
- 20 American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th edn (DMS–IV). Washington, DC: APA, 1994
- 21 Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin* 1979;**86**:420–8
- 22 Simon GE, Revicki D, von Korff M. Telephone assessment of depression severity. *Journal of Psychiatric Research* 1993;**27**: 247–52
- 23 Williams JB. A structured interview guide for the Hamilton Depression Rating Scale. *Archives of General Psychiatry* 1988;**45**: 742–7
- 24 Robinson DS, Rickels K. Concerns about clinical drug trials. *Journal of Clinical Psychopharmacology* 2000;**20**:593–6