# Assessment of depression diagnosis and severity: Comparability of telephone and in-person interview

*Christian Yavorsky, PhD* & Janet B.W. Williams, DSW**
** MedAvante, Inc.*

**Background:** Telephone-based assessment of depression has become more common in clinical and research contexts. Telephone assessment can provide more ready access to hard-to-reach populations and, for clinical trials, a more diverse sample. Despite these advantages, there are often questions about the tolerability of telephone interviews with depressed populations as well as the reliability of this administration mode. This review examines studies comparing the reliability and acceptability of telephone and in-person assessment of depression and seeks to address the following questions: Can telephone assessment of depression yield similar reliability and validity as the same assessment administered in-person? If so, are there differences in data quality, rapport and patient satisfaction? How can these answers be weighed against other factors like the often-prohibitive expense of doing large-scale projects with multiple follow-ups?

**Method:** Electronic databases including Medline, PsychINFO, Google scholar, Cochrane Database, CINAHL and the Wellcome Institute were searched using a list of key words. The key terms of 'telephone,' 'in-person,' 'face-to-face,' 'interview,' 'assessment,' 'personal,' 'depression' and 'psychiatric' were entered into the databases, along with adapted forms to ensure that a broad range of topics would be caught, e.g., 'tele*' which yielded articles about telemedicine, and 'psych*' which located psychiatric, psychometric and psychology. Terms such as 'rating,' 'structured interview,' 'validity,' and 'reliability' were also searched as were a number of common assessment tools, e.g., HAMD and MADRS. The time period covered 1967 until December of 2008.

A total of 84 articles were reviewed. The minimum criteria included the comparison of telephone and in-person interviews using a standard measure to assess a psychiatric condition. This was done so as not to inadvertently exclude articles that may have captured a depression rating, even if that was an incidental finding. Articles were then further narrowed to include only depression. Related articles were obtained by systematic review of the bibliographies of the above articles. Inclusion criteria were that the studies compared telephone and in-person assessment of depression. The first author screened the articles to ensure that the articles represented original research, conflicts of interest were reported, and appropriate controls were used.

**Results:** There were 84 articles reviewed with 14 meeting the narrower criterion for inclusion of comparing telephone to in-person assessment. Measures appropriate for comparison were reported in all studies, with intraclass correlation coefficients (ICC) reported in 36% (5/14), kappa reported in 43% (6/14), and 21% (3/14) used other methods to assess agreement, depending on the characteristics of the instrument studied. The range of ICCs (.80-.96) and range of kappas (.45-.96) demonstrate good agreement across multiple instrument types.

## Key findings:

**Agreement with in-person assessment**

In most cases there was good or excellent agreement between the two procedures.

**Economics**

Telephone assessments are a more economic way of accessing large patient populations as found by a number of large-scale studies including the STAR*D study.

**Assessing observed behaviors**

There were few or no changes in scores on key observable items, e.g., "apparent sadness" on the MADRS

**Patient satisfaction – Key findings**

Patients often prefer telephone assessment of depression and the literature suggests that this may be a function of the familiarity of the interface as well as the reduction in anxiety as compared to an in-person interview.

**Conclusion:** The aim of this review was to provide a summary and analysis of the literature comparing telephone to in-person interviews in depressed populations. The literature reviewed here suggests that for depression, telephone and in-person assessments are essentially equivalent, but adds the conditions that: using structured instruments to assess depression over the telephone with trained raters at appropriate intervals yields good reliability when compared to in-person interviews. Furthermore, telephone assessments are useful in a variety of contexts in which in-person interviewing may be difficult for geographic or economic reasons (e.g., large distances or large samples). Our overall conclusion is that rating depressive symptoms by telephone should be viewed as at least equivalent to in-person assessment, and in many instances is preferred by this patient group. However, a high level of rater training and the use of well-validated instruments improve agreement between modalities.

## DIAGNOSTIC SCALES

| Author | Year | Scale | Reported statistic | Value | N |
|---|---|---|---|---|---|
| Allen, K., Cull, A. & Sharpe, M. | 2003 | SCID | Not reported | NA | 219 |
| Cacciola, J.S., Alterman, A.I., Rutherford, M.J., McKay, J.R. & May, D.J. | 1999 | SCID | kappa | .64 (lifetime)<br>.66 (current) | 41 |
| Rohde, P., Lewinsohn, P. & Seeley, J. | 1997 | SCID | kappa | .96 (inter-rater)<br>.67 (test- retest) | 60 |
| Simon, G., Revicki, D. & Von Korff, M. | 1993 | SCID | ICC | .80 | 31 |
| Watson, C., Anderson, P., Thomas, D. & Nyberg, K. | 1992 | DIS | kappa | .60 (total)<br><.20 (depression) | 49 |
| Wells, K., Burnam, A., Leake, B. & Robins, L. | 1988 | DIS (lifetime depression) | kappa | .57 | 230 |

## SEVERITY SCALES

| Author | Year | Scale | Reported statistic | Value | N |
|---|---|---|---|---|---|
| Aneshensel, C.S., Frerichs, R.R., Clark, V.A. & Yokopenic, P.A. | 1982 | CES-D | Structural equation model | NA | 546 |
| Aneshensel, C.S. & Yokopenic, P.A. | 1985 | CES-D | Causal modeling (SEM) | NA | 546 |
| Burke, W.J., Roccaforte, W.H., Wengel, S.P., Conley, D.M. & Potter, J.F. | 1995 | GDRS | kappa | .52 | 101 |
| Feiger, A., Sikich, D. & Mason, R. | 2007 | MADRS<br>HAM-D | ICC<br>ICC | .93<br>.92 | 50 |
| Hermens, M., Ader, H., van Hout, H., Terluin, B., van Dyck, R. & Haan, M. | 2006 | MADRS | ICC | .65 | 66 |
| Kobak, K., Williams, J., Jeglic, E., Salvucci, D. & Sharp, I. | 2007 | MADRS | ICC | .93 | 35 |
| Pinto-Meza, A., Serrano-Blanco, A., Penarrubia, M., Blanco, E. & Haro, J. | 2005 | PHQ-9 | ICC | .94 | 346 |
| Potts, M., Daniels, M., Burnam, A, Wells, K. | 1990 | SI-HDRS | ICC | .92 | 498 |
| Simon, G., Revicki, D. & Von Korff, M. | 1993 | HAM-D<br>SCL-90 (depression) | ICC<br>ICC | .80<br>.90 | 31 |