

Reliability and Acceptability of Automated Telephone Surveys Among Spanish- and English-Speaking Mental Health Services Recipients

Benjamin B. Brodey,¹ Craig S. Rosen,² Inger S. Brodey,³
Breanne Sheetz,¹ and Jürgen Unutzer⁴

Interactive Voice Response (IVR), an automated system that administers surveys over the phone, is a potentially important technology for mental health services research. Although a number of studies have compared IVR to live interviews, few have looked at IVR in comparison to pencil-and-paper survey administration. Further, few studies have included subjects from those populations most likely to benefit from IVR technology, namely patients with lower education levels and non-English-speaking patients. This randomized clinical study, conducted at a community health center serving low-income English- and Spanish-speaking populations, assessed the reliability of an IVR-administered Brief Symptom Inventory (BSI) relative to a paper-and-pencil version. The study was adequately powered. Results showed that patients gave similar responses to the IVR and paper-and-pencil surveys; in addition, patients were generally equally satisfied with both experiences. We conclude that, while more large-scale research is needed, IVR can be a useful survey administration tool.

KEY WORDS: IVR; telephone; survey; BSI; satisfaction; CAT.

INTRODUCTION

Interactive Voice Response (IVR) is a relatively new survey administration technology that offers a number of advantages over more traditional forms of survey administration. In contrast to live or telephonic clinician interviews, IVR uses an automated system to deliver prompts and record touch-tone re-

sponses over the telephone. By integrating an algorithm for dynamically selecting question sequences and automating scoring and report generation, IVR can significantly reduce both the patient and administrative burdens of using surveys, while making the administration of a large number of surveys economically feasible. IVR surveys can also be easily modified to offer prompts in languages other than English, further simplifying the collection of patient-centered information in multilingual environments.

As interest in IVR technology for clinical surveys has grown, a number of studies have compared IVR favorably to its live counterparts, the live telephone interview and the clinician interview (e.g. Baer et al., 1995; Kobak, Greist, Jefferson, Mundt, & Katzelnick, 1999; Kobak et al., 1997; Mundt et al., 1998); fewer attempts have been made to compare IVR to paper-and-pencil survey results (e.g. Baer, Brown-Beasley, Sorce, & Henriques, 1993; Agel, Rocwood, Mundt, Greist, & Swiontkowski, 2001). The paper-IVR comparison is particularly important

¹TeleSage, Inc., P.O. Box 750, Chapel Hill, North Carolina 27514; e-mail: bbrodey@telesage.com and breannesheetz@hotmail.com.

²Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Palo Alto, Stanford, California 94305; e-mail: craig.rosen@stanford.edu.

³Curriculum in Comparative Literature, University of North Carolina, Chapel Hill, North Carolina 27599-3150; e-mail: ibrodey@telesage.com.

⁴Department of Psychiatry and Behavioral Sciences, University of Washington School of Medicine, P.O. Box 35650, Seattle, Washington 98195-5650; e-mail: unutzer@u.washington.edu.

⁵Correspondence should be directed to Benjamin B. Brodey, P.O. Box 750, Chapel Hill, North Carolina 27514.

since IVR is well suited to patients who do not read well and to non-English-speaking patients who may not have access to health care in their primary language.

This study explores the reliability of IVR in a community of both English and Spanish speakers by comparing IVR and paper-and-pencil administrations of the Brief Symptom Inventory (BSI) in Spanish and English. Unlike earlier studies, which had used primarily highly educated volunteers rather than the lower-income, less-educated patients most likely to benefit from the availability of IVR (Glass, 1997), this study focused on outpatients at a community health center serving low-income populations.

Our research objective was to determine whether results varied between paper-and-pencil and IVR survey administrations in English- and Spanish-speaking groups of subjects and to assess the satisfaction of subjects with each survey technology.

METHODS

The study was conducted in 1997 at a community healthcare provider serving a primarily low-income population composed of both Spanish and English speakers; the data analysis was conducted in 2004. Study participants were asked to take the Brief Symptom Inventory (BSI) three times at 1-week intervals; they were compensated \$5.00 for the first assessment and \$15.00 for the second and third survey administrations.

The BSI is an established 53-item self-report mental health instrument covering nine domains of psychiatric symptoms, yielding a single composite score called the Global Severity Index (GSI). After providing some basic demographic information, subjects were asked to consider their experience over the previous 30 days and assess the severity of each symptom described in the BSI on a 5-point scale ranging from "not at all" to "extremely." We chose the BSI for this study because it is a well-established instrument covering a broad range of symptoms and because it has been validated in English and Spanish. We used the written Spanish version validated by Acosta, Nguyen, and Yamamoto (1994) to record IVR voice prompts in Spanish for use with the SmartQ™ IVR survey software; in both the Spanish and English versions, subjects responded to questions using the telephone keypad. The system was configured to score the BSI and deliver reports in real time.

Study subjects were randomized into two groups to ensure that the order of survey administration would not affect the outcome. Group 1 took the IVR survey during the first administration, and the paper-and-pencil version during the second and third administrations. Group 2 took the paper-and-pencil version first and the IVR second and third. Three survey administrations allowed us to assess the test-retest reliability within each modality, as well as across modalities.

Test-retest reliability within the IVR and paper formats and correspondence of scores across the two different formats were assessed with Pearson correlations. The correlations were run for the entire sample, as well as for the English and Spanish subsamples. The significance of differences in correlation coefficients either within or across samples was assessed with Cohen's q (Cohen, 1988). Mean differences in scores obtained by IVR and paper formats were assessed using a repeated measures analysis of variance, including the between-subject main effect of language (English vs. Spanish), the within-subject effect of modality (paper vs. IVR), and the interaction of modality and language. Assuming scores obtained by paper and IVR formats are correlated at least .75, a sample of 100 subjects provides 80% power to detect within-subject differences of 0.2 SD , and 95% power to detect differences of 0.25 SD (Cohen, 1988).

After each survey administration, subjects completed four user satisfaction questions, responding on a 5-point Likert scale to questions about their ability to understand the questions, survey length, clarity of instructions, and their overall satisfaction with the survey. Higher responses indicated greater satisfaction. Mean differences in response to each of these questions in the IVR and paper modalities were assessed with a repeated measure analysis of variance, including the between-subject factor of language and the within-subject factors of modality and modality \times language.

RESULTS

The study was conducted at the Sea Mar Community Health Center in Seattle, Washington. The Health Center's primary mission is to provide health and human services to the low-income and seasonal farm worker communities. Approximately 60% of Sea Mar patients use Spanish as their primary language. All patients at the outpatient mental

health clinic were invited to participate in the study. Approximately 50% of those who visited the clinic during the study period elected to participate, yielding a final cohort of 107 subjects, 55 Spanish speaking and 52 English speaking, over the age of 18 years; mean age was 30 years among English-speaking subjects and 31 years among Spanish speakers. Sixty-four percent of subjects were women, and 87% were members of a racial or ethnic minority. The median household income of English-speaking participants was under \$10,000; median income of Spanish-speaking participants was between \$10,000 and \$20,000. Twenty-seven percent of English speakers had some post-secondary education, while 51% of Spanish speakers had some post-secondary education. By contrast, the 2003 Current Population Survey from the U.S. Census Bureau reports that 56.4% of non-Hispanic white Americans and 29.6% of Hispanics have some college education (U.S. Census Bureau, 2004).

Although the introductory material received by all participants stressed the subject's freedom to cease participation at any time during an individual survey administration without sacrificing compensation, every subject completed each survey he or she began. Attrition rates between sessions were also lower than anticipated. We had anticipated an attrition rate of 15–20%; the actual rate was 8%. Ninety-eight of the original 107 participants, 50 English speakers and 48 Spanish speakers, completed the study.

Overall, results did not demonstrate significant differences in subjects' response to IVR or paper-and-pencil surveys, or their satisfaction with each of the survey technologies. Subjects required the same amount of time to complete both paper and IVR surveys, approximately 10 minutes. Although GSI scores of English-speaking subjects were directionally higher than those of Spanish speakers, the repeated-measures ANOVA indicated this difference was not significant ($F_{1,96} = 1.1$, ns). Most importantly, there were no within-subject differences in scores by modality ($F_{1,96} = 0.0$, ns), nor was there a significant modality by language interaction ($F_{1,96} = 1.1$, ns).

Test-retest reliability of the BSI in the IVR format was $r = .95$ ($n = 45$), and BSI scores obtained by paper and IVR correlated $.91$ ($n = 98$). These are similar to the test-retest reliability previously reported for the standard paper version of the BSI (Derogatis & Melisaratos, 1983). Test-retest reliability of the IVR BSI did not differ significantly between

the participants who completed both measures in English ($r = .97$, $n = 23$) or in Spanish ($r = .88$, $n = 22$). Correspondence of scores across IVR and paper formats was also similar for English speakers ($r = .92$, $n = 48$) and Spanish speakers ($r = .90$, $n = 50$).

The only element of satisfaction which differed significantly by modality was clarity of instructions, which was rated more highly for paper than for IVR ($F_{1,96} = 7.7$, $p < .01$, $d = 0.35$). Clarity of instructions was rated more highly by English speakers than Spanish speakers ($F_{1,96} = 15.6$, $p < .001$), but the interaction of language \times modality was not significant ($F_{1,96} = 2.3$, ns).

Overall satisfaction was directionally higher among English speakers ($F_{1,96} = 2.8$, $p < .10$), but did not vary by modality or modality \times language ($F_{1,96} = 0.4$ – 0.6 , ns). Understanding of survey questions was rated more highly by English speakers than Spanish speakers ($F_{1,96} = 11.8$, $p < .001$), but did not differ by modality or modality \times language ($F_{1,96} = 0.3$ – 1.1 , ns). Ratings of survey length did not differ significantly by language, modality, or language \times modality ($F_{1,96} = 0.0$ – 1.6 , ns).

DISCUSSION

The results of this study are significant in that, unlike most previous IVR studies, they address the reliability and acceptability of IVR among the specific populations most likely to benefit from its widespread use. Our results show that IVR surveys generate patient responses similar to those generated by traditional paper surveys; differences between the two modalities were not statistically significant. Further, IVR proved to be as reliable as paper- and pencil-administered surveys upon retesting. Perhaps most importantly, patients reported similar satisfaction with IVR and with paper.

CONCLUSION

Patient-centered surveys are an increasingly important component of mental health care and mental health services research. As these instruments become more prevalent, economical methods must be developed to administer the large volumes of surveys required for research and clinical care. The development of Computer Adaptive Tests (CAT), a key component in patient-centered outcomes measurement (McHorney, 2003), offers similar

challenges for researchers and practitioners. This study demonstrates that IVR may well prove to be such a method. Although additional attention should be given to clarifying IVR instructions for both English and Spanish speakers, IVR survey administration has valuable applications in research and health outcomes tracking. Further large-scale research into the applicability of IVR technology in mental health services research is needed.

ACKNOWLEDGMENTS

This research was supported by a National Institutes of Mental Health SBIR grant (1 R43 MH57614-01A1) entitled "New Automated Telephone Technology for Mental Health." The study received a Single Project Assurance for research with human subjects from the National Institutes of Mental Health, and was approved by the board of directors at Sea Mar Community Health Centers in Seattle, Washington. The authors acknowledge the assistance of Keith Claypoole, Ph.D., Department of Psychology, University of Hawai'i at Manoa, and Wayne J. Katon, M.D., Department of Psychiatry and Behavioral Sciences, University of Washington School of Medicine, in developing the methodology for this study.

REFERENCES

- Acosta, F. X., Nguyen, L. H., & Yamamoto, J. (1994). Using the Brief Symptom Inventory to profile monolingual Spanish-speaking psychiatric outpatients. *Journal of Clinical Psychology, 50*, 723-726.
- Agel, J., Rocwood, T., Mundt, J. C., Greist, J. H., & Swiontkowski, M. (2001). Comparison of interactive voice response and written self-administered patient surveys for clinical research. *Orthopedics, 24*, 1155-1157.
- Baer, L., Brown-Beasley, M. W., Sorce, J., & Henriques, A. I. (1993). Computer-assisted telephone administration of a structured interview for obsessive-compulsive disorder. *American Journal of Psychiatry, 150*, 1737-1738.
- Baer, L., Jacobs, D. G., Cukor, P., O'Laughlen, J., Coyle, J. T., & Magruder, K. M. (1995). Automated telephone-screening survey for depression. *Journal of the American Medical Association, 274*, 1943-1944.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale, NJ: Earlbaum.
- Derogatis, L. R., & Melisaratos, N. (1983). The Brief Symptom Inventory: An introductory report. *Psychological Medicine, 13*, 595-605.
- Glass, R. M. (1997). Minds and machines: Interactive voice response technology to detect mental disorders. *Journal of the American Medical Association, 278*, 946.
- Kobak, K. A., Greist, J. H., Jefferson, J. W., Mundt, J. C., & Katelnick, D. J. (1999). Computerized assessment of depression and anxiety over the telephone using interactive voice response. *M.D. Computing, 16*, 64-68.
- Kobak, K. A., Taylor, L. H., Dottle, S. L., Greist, J. H., Jefferson, J. W., Burroughs, D., Mantle, J. M., Katelnick, D. J., Norton, R., Henk, H. J., & Serlin, R. C. (1997). A computer-administered telephone interview to identify mental disorders. *Journal of the American Medical Association, 278*, 905-910.
- McHorney, C. A. (2003). Ten recommendations for advancing patient-centered outcomes measurement for older persons. *Annals of Internal Medicine, 139*, 403-409.
- Mundt, J. C., Kobak, K. A., Taylor, L. V. H., Mantle, J. M., Jefferson, J. W., Katelnick, D. J., & Greist, J. H. (1998). Administration of the Hamilton Depression Rating Scale using interactive voice response technology. *M.D. Computing, 15*, 31-39.
- U.S. Census Bureau. (2004). *Educational attainment in the United States: 2003*. Retrieved September 8, 2004, from <http://www.census.gov/prod/2004pubs/p20-550.pdf>.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.