

## Interactive voice response: Review of studies 1989–2000

ROSS CORKREY and LYNNE PARKINSON

*University of Newcastle, New South Wales, Australia*

A systematic review of the use of interactive voice response (IVR) was conducted. IVR is a telephone interviewing technique in which the human speaker is replaced by a high-quality recorded interactive script to which the respondent provides answers by pressing the keys of a touch telephone (touchphone). IVR has numerous advantages, including economy, autonomy, confidentiality, access to certain population groups, improved data quality, standardized interviewing, multilingual interfaces, and detailed longitudinal assessments. Despite this, there have been few applications of IVR. Previous studies have been in the areas of information services, reminder calls, monitoring, assessment, experimentation, interventions, and surveys. Areas that have received little attention have been the systematic evaluation of voice, multilingual interfaces, touchphone prevalence, survey response rates, use by the elderly, and acceptability.

Numerous methods have been used to collect survey data, including face-to-face, self-administered, mail, and telephone modes. Interactive voice response (IVR), also known as touch-tone data entry, is a relatively recent development in telephone surveying, which has yet to be formally tested in population surveys. In this paper, we will discuss the range of uses IVR has had to date, critically review the work done so far, and suggest where further work may be required, particularly in respect to population surveys. We consider this an essential first step in understanding a technology that may have an exciting potential for survey research.

In IVR, the human speaker is replaced by a high-quality recorded interactive script to which the respondent provides answers by pressing the keys of a touch telephone (touchphone). It differs from a computer-assisted telephone interview (CATI) by lacking an interviewer to read questions and enter the answers into the computer. Many attributes of CATI are shared by IVR, including those listed by Nicholls (1988): automatic logical skipping or branching based on answers to earlier questions; randomized option and question order; interviews that can be interrupted and resumed at a later time; allowing single-choice, as well as open-ended, responses to questions; validating responses as they are entered; providing feedback on inappropriate responses; call scheduling; and automatic interview and recording keeping.

However, IVR can offer considerable advantages: An IVR system can run automatically and continuously; re-

spondents may proceed through an interview at their own rate; with the involvement of field workers and the use of cellular telephones, IVR can allow access to hard-to-reach groups, such as sex industry workers, drug users, and homeless people (e.g., Alemagno, Cochran, et al., 1996); there is no interviewer bias, and interrater reliability is irrelevant; there is an opportunity for interview standardization through the use of high-quality recorded voices and standardized pronunciation; since IVR uses a standardized interface and any assessment is conducted by a computer, interview accuracy is improved; the cost per interview is likely to be small (Havice, 1989); more precise and detailed longitudinal assessments should be possible (Mundt, Perrine, Searles, & Walter, 1995; Mundt, Searles, Perrine, & Helzer, 1995); and its greater confidentiality may result in a lower response bias (Tourangeau & Smith, 1996).

Other recent methods include Web and e-mail surveys that may share most of the above attributes. However, these are currently limited by low response rates (Sheehan & Hoy, 1999) and poor frame coverage for general populations (Couper, 2000), although a mixed-mode approach may reduce coverage error (Schaefer & Dillman, 1998).

There is the potential to conduct automated screening of large populations in order to identify particular target groups that may then be followed in more detail once informed consent has been given. Where appropriate, such an application may also require procedures with which to handle identification of false positive cases. The complete anonymity and increased perception of confidentiality provided by IVR may also have particular benefits when used to assess sensitive behaviors, since more confidential interviewing methods have been shown to produce greater reporting levels of sensitive behaviors (Tourangeau & Smith, 1996).

An IVR system typically involves running software on an ordinary personal computer in which a voice card has

---

This project was funded by a grant from Hunter Medical Research Institute and forms part of the doctoral studies of R.C., which is supported by an Australian Postgraduate Scholarship. Correspondence concerning this article should be addressed to R. Corkrey, P. O. Box 491, Wallsend NSW 2287, Australia (e-mail: corkrey@optusnet.com.au).

been installed. The voice card is connected to one or more telephone lines and allows calls to be made or received entirely automatically, plays or records speech, and responds to keys pressed. The computer may ask questions of the caller by playing sound files that contain recorded questions. The caller responds by pressing the keys of their touchphone.

This paper critically reviews the literature on IVR applications from 1989 to 2000. Application areas and factors that may affect its acceptability, reliability, and feasibility, particularly in respect to its potential for population surveys and health research, will be examined.

## METHOD

The business, medical, and statistical literature was examined for IVR studies by searching five on-line databases (American Psychological Association, 2001; American Statistical Association & the Institute of Mathematical Statistics, 2001; Information Access Co., 2001; National Library of Medicine, 2001; Sociological Abstracts, 2001). The following search terms were used: interactive voice response, IVR, computer assisted telephone interview, CATI, interview mode, mode effect, data grabber, touch tone, voice response unit, and VRU.

Abstracts of references located by the search were examined, and only those making reference to IVR or a synonym were retained. Articles discussing the use of speech recognition, technological descriptions, voice mail studies, or marketing applications were discarded. As is usual in systematic reviews (Chalmers et al., 1987), only peer-reviewed journals were included, since other sources may be inconsistent in content and accessibility.

Full copies of all relevant articles were obtained. Bibliographies were searched for further references, and

these were treated as described above. The process was terminated when no more articles were identified. All the articles were coded using the following headings: study area, touchphone prevalence, voice assessment, number of voices, acceptability, nonresponse rate, sampling method, sample size, and sample attributes (headings are defined in the Results and Discussion section).

The articles were grouped into the following study areas: information services, reminder calls, patient monitoring, consumption monitoring, psychological assessment and treatment, psychological experiments, behavioral interventions, and surveys. Areas that need further attention will be discussed. A total of 54 articles were reviewed and are summarized in Appendix A.

## RESULTS AND DISCUSSION

### Use of IVR Over Time

The reviewed studies described IVR applications from 1989 onward, and we are not aware of any earlier work on this subject in refereed journals, except for a technical description of a clinical trial randomization system (Cheung, Entine, & Klotz, 1977). Since 1994, the rate of publication of IVR articles has modestly increased, whereas CATI papers have been published since 1978 (see Figure 1). The relative scarcity of IVR applications probably results from its more recent development, and it seems likely that the number of IVR applications will increase as the technology matures.

### Study Areas

In the reviewed studies, IVR has primarily been used in health areas—for example, alcohol and drug monitoring—and for survey applications to nonhealth areas (see Appendix A).

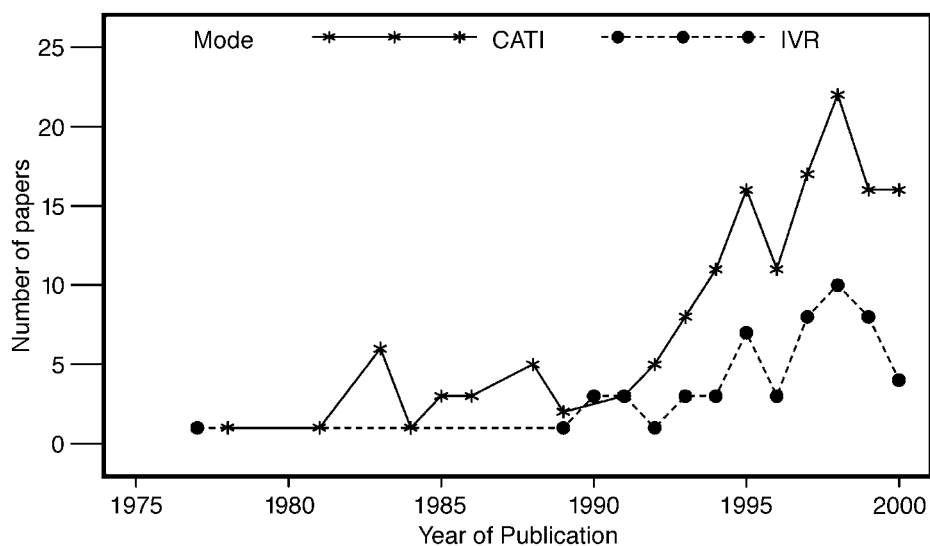


Figure 1. Number of interactive voice response (IVR) and computer-assisted telephone interview (CATI) publications per year (1977–2000).

**Information services.** IVR systems can receive inbound calls in order to supply information, including educative services, to screen for health risks, and possibly to conduct behavioral interventions. The confidential nature of an IVR system also lends itself to applications in which assessments or diagnostic test results on sensitive topics may be provided.

Two smoking cessation information lines and four screening services (depression, breast cancer, HIV, and substance abuse) that used advertising, relied on volunteers, or used convenience samples were studied (Baer et al., 1995; Burke, 1993; Dyches, Alemagno, Llorens, & Butts, 1999; Havel & Wright, 1997; Kadison, Pelletier, Mounib, Oppedisano, & Poteat, 1998; Schneider, Schwartz, & Fast, 1995).

One smoking cessation service (Burke, 1993) received 5,345 callers, with 78% listening to the full message, but there was no assessment of efficacy. In another smoking cessation study (Schneider et al., 1995), applications were received from 975 individuals, of whom 571 used the system. The overall quit rate was 35%, but 45% of those who called at least five times quit.

In a depression screening service (Baer et al., 1995), 1,812 calls were received from university students and corporate employees, with 88% listening to the whole message; 26% were identified as at least moderately depressed, and 75% found the service helpful. In a breast cancer risk screening service (Kadison et al., 1998), 343 calls were received from a corporate population of 8,900, with 145 identified as having at least one risk factor; in a follow-up interview of 189 women, 70 (52%) reported that they had changed their screening habits.

A hepatitis B risk screening service (Havel & Wright, 1997) received 558 calls from a university community of 80,000 people, with 85% listening to the entire message, 299 identified as at risk, and 124 subsequently vaccinated. The last study (Dyches et al., 1999) screened 724 primary care patients for substance abuse and obtained a 42% compliance rate.

In three studies, preexisting information services were provided to known populations (Frank et al., 1997; Kempe, Dempsey, & Poole, 1999; Mahoney, Tennstedt, Friedman, & Heeren, 1999). A pediatric clinic information line (Kempe et al., 1999), supported by mail booklets and posters, received 1,100 calls from a patient population of 8,365, with 86% of a follow-up sample reporting satisfaction and 69% reporting that the call had made a visit to the clinic unnecessary. A geriatric functional status system (Mahoney et al., 1999) assessed 20 disabled elders and obtained results similar to those with a telephone interview but identified fewer functional disabilities than were identified with a personal visit. In the last study (Frank et al., 1997), 1,147 subjects used HIV home collection kits and obtained the result by ringing an IVR system. When offered counseling, 1,070 (85%) chose an IVR session, rather than a telephone counselor.

In a pilot study (Alemagno, Frank, Mosavel, & Butts, 1998), IVR was used to screen 116 adolescents for alco-

hol and drug consumption, sexual activity, stress, depression, and suicidal thoughts. The accuracy of the self-reported risks was assessed by 22 physicians, resulting in an overall agreement of 84%.

Except for three studies (Alemagno et al., 1998; Frank et al., 1997; Schneider et al., 1995) in which prior registration was used, the number of calls could not be taken to represent the number of unique users. All the studies relied on volunteer samples, implying that their respective sample characteristics cannot be taken to represent the populations from which they were taken.

**Reminder calls.** In seven studies (Dini, Linkins, & Chaney, 1995; Leirer, Morrow, Tanke, & Pariente, 1991; Lieu, Capra, Makol, Black, & Shinefield, 1998; Linkins, Dini, Watson, & Patriarca, 1994; Stehr-Green, Dini, Lindgren, & Patriarca, 1993; Tanke & Leirer, 1994; Tanke, Martinez, & Leirer, 1997), it has been investigated whether automated reminder calls can be successful in increasing appointments and treatment compliance.

In three of four studies, childhood immunization rates were increased by IVR reminder calls. In one study (Stehr-Green et al., 1993), the immunization rate for 112 children assigned to an intervention group and 110 children in a control group did not differ significantly. In another study (Linkins et al., 1994) of 4,636 children listed in an immunization database, increases were obtained in preschool immunization rates (36% vs. 28%,  $p < .01$ ). In another (Dini et al., 1995), increases were obtained in immunization rates for 277 child patients at public health clinics (20% vs. 49%,  $p < .05$ ). In the last study (Lieu et al., 1998) of 4,636 children listed in an immunization database, increases were obtained in preschool immunization rates after the children received a combination of an IVR call and a letter (58%), rather than just an IVR call (44%) or a letter (44%).

Increased treatment compliance was obtained by using IVR reminder calls in three studies (Leirer et al., 1991; Tanke & Leirer, 1994; Tanke et al., 1997). Appointment rates for 2,008 tuberculosis care patients in public clinics increased from 52% to 62% ( $p < .01$ ; Tanke & Leirer, 1994). In another study (Tanke et al., 1997) of 701 patients, tuberculin skin test return failures were reduced from 14% to 7% ( $p < .05$ ). In the last article (Leirer et al., 1991), improved medication adherence was obtained with 16 elders by reducing complete forgettings to 10.6%, as compared with 41.1% in a control group.

**Patient monitoring.** Both inbound and outbound dialing has been used to monitor patients. However, all the studies were uncontrolled, and so little can be concluded on the efficacy of such methods.

In four studies (Albisser, Harris, Sakkal, Parson, & En Chao, 1996; Alemi et al., 1994; Harding et al., 1997; Meneghini, Albisser, Goldberg, & Mintz, 1998), patients called into an IVR system in order to report symptoms or progress.

An irritable bowel syndrome study (Harding et al., 1997) monitored the symptoms of 640 patients and showed a compliance rate of between 74% and 83%. However, in

another study (Alemi et al., 1994) of 42 recovering drug users and alcoholics, compliance was reduced from 100% to 20% after 19 weeks. Despite the attrition, the IVR compliance always exceeded written questionnaire completion rates. In a study (Meneghini et al., 1998) of 184 diabetic patients, blood glucose reports were obtained an average of 58 times per month, and average HbA1c was reduced by 0.9% after 12 months ( $p = .04$ ). Another study of 204 diabetic patients (Albisser et al., 1996) also showed a reduction of HbA1c of 1.0%–1.3% ( $p < .05$ ).

Outbound dialing was used to directly contact patients in four studies (Patel & Babbs, 1992; Piette, 1999; Piette & Mah, 1997; Piette, McPhee, Weinberger, Mah, & Kraemer, 1999). Two studies (Piette, 1999; Piette et al., 1999) used a bilingual interface (English and Spanish). The first (Piette et al., 1999) monitored 252 diabetic patients over 12 months and obtained a completion rate of 69%. The second (Piette, 1999) monitored 226 English-speaking and 30 Spanish-speaking diabetic patients. The third (Piette & Mah, 1997) and fourth (Patel & Babbs, 1992) studies of 65 diabetic and 2 cardiac patients demonstrated technical aspects of the systems.

**Consumption monitoring.** IVR was used to obtain periodic alcohol and drug use reports in seven articles (Bardone, Krahn, Goodman, & Searles, 2000; Mundt, Perrine, et al., 1995; Mundt, Searles, et al., 1995; Perrine, Mundt, Searles, & Lester, 1995; Searles, Helzer, & Walter, 2000; Searles, Perrine, Mundt, & Helzer, 1995), resulting in detailed longitudinal data and higher self-report rates than in less confidential or frequent methods.

A 28-day study (Perrine et al., 1995) of 30 males obtained a good correlation between blood alcohol concentration and IVR self-report ( $r = .72$ ) and collateral report ( $r = .89$ ). A study (Searles et al., 1995) of 55 couples with a drinking male obtained 93% compliance and IVR self-report rates that were consistently higher than those in which the quantity–frequency method (QF) was used.

A 12-week study (Bardone et al., 2000) of binge eating and drinking among female college students obtained a compliance of 84%. Alcohol IVR self-report agreed with the timeline follow-back method (TLFB; Sobell & Sobell, 1992), but binge eating IVR self-report was 2.2 times greater than the TLFB ( $p < .01$ ).

In a 366-day study (Searles et al., 2000) of 33 males in which daily IVR alcohol consumption report was compared with 13 weekly TLFB reports, TLFB underreported heavy drinking days, as compared with the IVR self-report. In a 16-week study (Mundt, Perrine, et al., 1995; Mundt, Searles, et al., 1995) of 55 males, daily report compliance was 93%. IVR self-report rates were consistently higher than those from QF, and dependent drinkers were more likely to drink at regular 7-day intervals, whereas non-dependent drinkers had longer intervals.

**Psychological assessment and treatment.** IVR was used to assess psychological states (Baer, Brown-Beasley, Sorce, & Henriques, 1993; Kobak, Greist, Jefferson, Mundt, & Katzelnick, 1999; Kobak, Taylor, Dottl, Greist,

Jefferson, Burroughs, Katzelnick, & Mandell, 1997; Kobak, Taylor, Dottl, Greist, Jefferson, Burroughs, Mantle, et al., 1997; Mundt et al., 1998) and provide self-treatment (Bachofen et al., 1999; Greist et al., 1998; Marks et al., 1998; Osgood-Hynes et al., 1998). Since all the studies used clinic patients or remunerated volunteers, its use with community samples remains untested.

In one study (Baer et al., 1993) of 18 patients, paper and clinician-administered versions of the Yale-Brown Obsessive-Compulsive scale for obsessive-compulsive disorder (OCD) were highly correlated with IVR ( $r = .97$  and  $.99$ ), and the mean scores did not differ.

In a study (Kobak et al., 1999) of depression and anxiety among 192 psychiatric patients, rates obtained by IVR and clinicians were almost identical, were correlated ( $r = .96$  and  $.65$ ), and had similar psychometric properties.

In a study (Mundt et al., 1998) of 367 clinic patients and remunerated community members, rates of depression obtained by using IVR and clinicians were correlated and were equivalent after conversion to Z statistics.

In a study (Kobak, Taylor, Dottl, Greist, Jefferson, Burroughs, Katzelnick, & Mandell, 1997; Spitzer et al., 1994) of 51 outpatients, IVR, computers, and clinicians showed similar rates of psychiatric disorders, using the PRIME-MD. In another study (Kobak, Taylor, Dottl, Greist, Jefferson, Burroughs, Mantle, et al., 1997) that used the PRIME-MD with 200 primary care patients, the rate of alcohol abuse was significantly greater with IVR than with clinicians (15% vs. 7.5%;  $p < .05$ ).

An IVR self-treatment system for OCD was examined in two uncontrolled studies (Greist et al., 1998; Marks et al., 1998) of 63 OCD patients, which demonstrated that subjects could self-assess, and another (Bachofen et al., 1999) with 21 OCD patients, which obtained improvement in 10 who completed two or more sessions ( $p < .01$ ).

In an uncontrolled study (Osgood-Hynes et al., 1998) of 41 patients, rates of depression declined after self-treatment using a program involving videotapes, booklets and IVR being used to assess depression and to provide self-help aids ( $p < .01$ ).

**Psychological experiments.** Two studies have used IVR as an experimental aid (Mundt, Kelleher, Perrine, & Searles, 1997; Temple et al., 1998). In the first study (Mundt et al., 1997), IVR was used for psychological assessment. Ten subjects successfully performed psychomotor and cognitive functioning tests, and the performances of another 7 subjects were monitored after they had consumed alcohol. In the second study (Temple et al., 1998), IVR was used as a confidential reporting method to assess concordance between 26 pairs of surgeons and patients following a postbiopsy consultation; only 50% agreed about patient understanding and 27% about thoroughness of discussion.

**Behavioral interventions.** IVR was used to conduct inexpensive and automatic behavioral interventions in five studies (Ershoff et al., 1999; Hyman, Herd, Ho, Dunn,

& Gregory, 1996; Hyman, Ho, Dunn, & Simons-Morton, 1998; Piette, Weinberger, & McPhee, 2000; Piette, Weinberger, McPhee, et al., 2000), but only two articles (Piette, Weinberger, & McPhee, 2000; Piette, Weinberger, McPhee, et al., 2000) in which personal contact had been used showed significant improvement.

In a randomized control trial (RCT; Hyman et al., 1996) of a cholesterol-lowering diet protocol involving 115 subjects, 59 received fortnightly IVR calls, and the remainder had no further contact. No difference between groups was detected after 6 months ( $p = .36$ ). In another RCT (Hyman et al., 1998) of a cholesterol-lowering diet protocol involving 123 subjects, 65 received IVR calls and questionnaires, and all had weekly classes. No significant difference was observed between the two groups ( $p = .58$ ).

In a randomized trial (Ershoff et al., 1999) of smoking cessation, 390 women received booklets and either called into an IVR system or received motivational interviewing. Compliance with the IVR method was only 21%, and no significant differences were detected ( $p > .05$ ).

In an RCT (Piette, Weinberger, & McPhee, 2000; Piette, Weinberger, McPhee, et al., 2000) that involved 280 English and Spanish speakers, the intervention consisted of fortnightly bilingual IVR calls, with telephone follow-up by a nurse. After 12 months of intervention, the subjects had better glycemic control than did control subjects (44% vs. 31%;  $p = .04$ ; Piette, Weinberger, McPhee, et al., 2000), had fewer symptoms of depression ( $p = .02$ ; Piette, Weinberger, & McPhee, 2000), and were more self-efficacious (Piette, Weinberger, & McPhee, 2000).

**Surveys.** There have been a few applications of IVR to surveys (Alemagno et al., 1996; Havice, 1989, 1990a, 1990b; Havice & Banks, 1991; Phipps & Tupek, 1991; Troutman, Murray, & Norlander, 1990), but response rates have generally been low.

Three surveys (Havice, 1989) on media use, two ( $n = 1,309$  and  $4,062$ ) in which interviewers were used and one ( $n = 709$ ) in which an IVR survey was used, resulted in completion rates of 55%, 74%, and 7.5%, respectively. An IVR survey (Troutman et al., 1990) of accidental domestic poisoning ( $n = 33,458$ ) resulted in 6,478 answered calls and 2,703 completions. The poisoning rate was estimated as 109 per 1,000, which was higher than in other studies quoted by the authors (Troutman et al., 1990), possibly owing to confidential self-report.

In an IVR survey (Havice, 1990a) of 877 numbers, 424 numbers were eligible, and 29 completions were obtained. In 224 follow-up interviews of nonrespondents, only 11% reported having refused because it was a computer interview. In an IVR survey (Havice, 1990b) of student housing conditions, 1,228 numbers were attempted, resulting in 380 partial completions and 224 completions. In a media survey (Havice & Banks, 1991) resulting in 485 IVR and 500 CATI interviews, the IVR method obtained fewer respondents who were at least 50 years of age, but no other demographic differences between IVR and CATI were obtained.

In an IVR survey (Phipps & Tupek, 1991) of 465 businesses, all responded, and there was little item nonresponse. In a study (Alemagno et al., 1996) combining IVR with mobile phones, 207 homeless people were interviewed on substance abuse, and their self-reports were compared with a biochemical assay, which resulted in a sensitivity of 69% and a specificity of 91%.

### Additional Areas for Research

**Voice.** One obvious aspect of interest to IVR research is the choice of voice. The voice is the only means by which an IVR system can engage the subject's interest sufficiently long to meet a study's objectives. Despite the importance of voice, only two studies (Baer et al., 1995; Phipps & Tupek, 1991) assessed voices. One study (Baer et al., 1995) used focus groups, and the other (Phipps & Tupek, 1991) used a follow-up CATI.

Most of the studies did not report how many speakers were used, but in 15 articles (Baer et al., 1995; Bardone et al., 2000; Havice, 1989, 1990b; Havice & Banks, 1991; Leirer et al., 1991; Meneghini et al., 1998; Piette & Mah, 1997; Piette, Weinberger, & McPhee, 2000; Piette, Weinberger, McPhee, et al., 2000; Schneider et al., 1995; Stehr-Green et al., 1993; Tanke & Leirer, 1994; Tanke et al., 1997; Troutman et al., 1990), a single voice was used.

Various voice types were used: (1) alternating male and female voices (Baer et al., 1995), reportedly to maintain the subjects' interest, (2) a synthetic voice (Meneghini et al., 1998), and (3) multiple female voices speaking in different languages (Tanke & Leirer, 1994).

Most of the studies lacked consideration of standardization of voices, voice characterization, the numbers or speakers within scripts, or the choice of voice, despite work in the survey literature (see Appendix B) indicating that the voice quality of interviewers can affect refusal rates (Oksenberg & Cannell, 1988; Oksenberg, Coleman, & Cannell, 1986) and the perception of credibility (Gélinas-Chebat & Chebat, 1996). Self-report rates can be influenced by interviewers dropping or raising the tone of their voices at the end of questions (Barath & Cannell, 1976). A faster speaking pace may be perceived as being more persuasive (Miller, Maruyama, Beaber, & Valone, 1976; Smith & Shaffer, 1991) than a slow pace, and a soft speaker may be perceived as less assured than a speaker with a high volume (Page & Balloun, 1978).

A closer attention to voice would have two benefits: (1) standard assessments of voices with suitable scales (Laver, 1968; Oksenberg et al., 1986) would allow greater comparability of studies; (2) IVR may also allow the effect of voice type on response rate to be systematically explored in a population survey context. Since existing scales (e.g., Oksenberg et al., 1986) may not be practical owing to their complexity, the design of new scales may be a suitable first step.

**Multilingual interfaces.** One area in which IVR may be especially useful is economical provision of services to minority language groups that may otherwise be excluded. For example, Dyches et al. (1999) found that

18% of nonrespondents refused because of language. Despite this, multilingual IVR is a little-explored area.

All the studies used English messages, but some included additional languages: Spanish (Frank et al., 1997; Lieu et al., 1998; Piette, 1999; Piette et al., 1999; Piette, Weinberger, & McPhee, 2000; Piette, Weinberger, McPhee, et al., 2000; Tanke & Leirer, 1994; Tanke et al., 1997), Vietnamese (Tanke & Leirer, 1994; Tanke et al., 1997), Cantonese (Lieu et al., 1998), and Tagalog (Tanke & Leirer, 1994).

Three studies reported IVR usage by language groups. One (Piette et al., 1999) showed that Spanish and English speakers used the system equally, and in another (Piette, 1999), Spanish speakers were more likely to listen to optional self-care messages than were English speakers (64% vs. 36%;  $p < .0001$ ). In the third (Lieu et al., 1998), non-English interface was selected by 6.7% of the subjects.

**Touchphones.** Since IVR usually requires the subject to have access to a touchphone, the prevalence of touchphones is of particular interest. Not all touchphones may be acceptable, since some may not be appropriately configured. For example, in one study (Hyman et al., 1998) in which 88% of the respondents had touchphones, approximately one third were unsuitable for use with IVR.

Where touchphones are common, an IVR system may provide a useful service, but it may not be feasible in low touchphone prevalence areas. In Australia, nonownership of telephones is associated with households with lower incomes, the elderly, or young adults (Steel, Vella, & Harrington, 1996). It may be speculated that touchphones have similar socioeconomic associations.

Two studies provided anecdotal estimates of touchphone prevalence of 94.8% (Havice, 1990a) and 75%–98% (Havice & Banks, 1991). It must be concluded that touchphone prevalence and its relationship to the demographic composition of populations are still not known. This topic is further examined elsewhere in this issue (Corkrey & Parkinson, 2002).

**Survey response rate.** Response rate is important in survey applications (Platek & Gray, 1986). Even if a random sampling method is used, obtaining a very low response rate can produce the equivalent of a volunteer sample. The effect is to widen the uncertainties of estimates in proportion to degree of nonresponse (Cochran, 1977). For IVR to be considered a viable survey method, an estimate of the expected response rate is needed.

Two survey studies (Alemagno et al., 1996; Havice & Banks, 1991) did not provide useful estimates. The first (Havice & Banks, 1991) did not report a response rate or the number of attempted contacts, and the second (Alemagno et al., 1996) used a volunteer sample for which a response rate is meaningless.

Response rates were available in five studies (Havice, 1989, 1990a, 1990b; Phipps & Tupek, 1991; Troutman et al., 1990). Response rates were 7.5% (Havice, 1989), 6.8% (Havice, 1990a), 18.2% (Havice, 1990b), and 58% (Troutman et al., 1990). The last study (Phipps & Tupek,

1991) showed a complete response, but this was an establishment survey, which did not provide useful information for other types of surveys.

The response rate for IVR surveys appears modest. This could be explained by the lack of any means of retaining respondents once contacted. This possibility is explored elsewhere in this issue (Corkrey & Parkinson, 2002).

**Sample characteristics.** Sample bias is another issue relevant to IVR surveys. Bias in demographic variables cannot be ascertained from many studies owing to small sample sizes or nonrandom sampling. The median number of participants was 204, the smallest was 3, and the largest 8,002 (see Appendix C).

Many of the studies also had samples with unusual demographic compositions, which suited the studies' purposes but did not aid in examining applicability of IVR to surveys. For example, some restricted their samples to all-male (Mundt, Perrine, et al., 1995; Mundt, Searles, et al., 1995; Searles et al., 2000; Searles et al., 1995), all-female (Kadison et al., 1998), or remunerated subjects (Alemagno et al., 1996; Alemi et al., 1994; Bardone et al., 2000; Frank et al., 1997; Kobak, Taylor, Dottl, Greist, Jefferson, Burroughs, Katzelnick, & Mandell, 1997; Kobak, Taylor, Dottl, Greist, Jefferson, Burroughs, Mantle, et al., 1997; Leirer et al., 1991; Mundt et al., 1997; Mundt, Perrine, et al., 1995; Mundt, Searles, et al., 1995; Perrine et al., 1995; Searles et al., 2000; Searles et al., 1995), or to subjects reused from other studies (Kobak, Taylor, Dottl, Greist, Jefferson, Burroughs, Mantle, et al., 1997; Mundt et al., 1997; Mundt et al., 1998; Mundt, Perrine, et al., 1995; Mundt, Searles, et al., 1995; Searles et al., 2000; Searles et al., 1995).

Some studies do provide indications of sample bias. In a comparison of IVR and CATI (Havice & Banks, 1991), both mean age and number of years of education were reported as being significantly less when IVR was used than when CATI was used. However, since the response rate was not reported, these comparisons are not very useful. In another comparison of IVR and CATI surveys (Havice, 1989), IVR self-reported age and sex were comparable those in two CATI surveys, but there was no statistical test provided. In addition, the IVR and CATI surveys covered different topics, and ages were recorded differently, making comparisons difficult. Demographic data from a IVR survey (Havice, 1990b) of university students were reported as being comparable to data from official university records, but since the self-reported ages were compared with ages calculated using years enrolled, this comparison is unreliable. However, the gender ratio did not appear to differ significantly within the limits of the sample size ( $n = 224$ ).

Given these limitations, the reviewed studies can give little or no indication of the demographic biases that may be expected in IVR surveys. This possibility is explored elsewhere in this issue (Corkrey & Parkinson, 2002).

**Elderly respondents.** The respondent's age is of particular interest, since older respondents may have greater

difficulty in using an IVR interface or may be less likely to respond to an IVR survey.

One study (Linkins et al., 1994) reported that the demographic profile of respondents did not differ between IVR and a control group. Another (Tanke & Leirer, 1994) found no significant effect owing to age or sex in the effectiveness of an IVR appointment reminder system. However, one study (Piette & Mah, 1997) reported descriptive data that suggested that older subjects were more likely to listen to at least two health promotion IVR messages. Another (Albisser et al., 1996) reported, without any statistical test, that older subjects were slower to learn the use of the IVR system.

In experiments of memory recall in which recorded messages were used, older subjects had poorer memory recall than did younger subjects, but this difference could be eliminated with a single imposed repetition (Morrow, Leirer, Carver, Tanke, & McNally, 1999b). Age-related memory recall differences could also be reduced when recorded messages were designed to be compatible with memory schemas (Morrow, Carver, Leirer, & Tanke, 2000). However, age differences were not reduced when message repetition was optional (Morrow, Leirer, Carver, Tanke, & McNally, 1999a).

These studies suggest that IVR interfaces should repeat messages when the elderly are dealt with and that messages should be organized according to memory schemas, which can be determined by using focus groups.

**Acceptability.** The use of IVR is likely to be novel to most people, and therefore, its acceptability needs to be carefully assessed. IVR technology is increasingly used by individuals for financial management (Australian Bureau of Statistics, 1998), but this does not imply that it would be acceptable in, for example, household drug surveys.

In only 18 studies (Alemi et al., 1994; Baer et al., 1993; Baer et al., 1995; Burke, 1993; Ershoff et al., 1999; Greist et al., 1998; Harding et al., 1997; Kadison et al., 1998; Kempe et al., 1999; Kobak et al., 1999; Kobak, Taylor, Dottl, Greist, Jefferson, Burroughs, Katzelnick, & Mandell, 1997; Lieu et al., 1998; Osgood-Hynes et al., 1998; Phipps & Tupek, 1991; Piette & Mah, 1997; Tanke & Leirer, 1994; Tanke et al., 1997; Temple et al., 1998) have researchers attempted to measure acceptability, and there was little agreement on the methods used.

Three studies (Baer et al., 1995; Kobak et al., 1999; Kobak, Taylor, Dottl, Greist, Jefferson, Burroughs, Katzelnick, & Mandell, 1997) showed that IVR was less preferred than alternative methods (face-to-face or CATI), but two of the studies (Kobak et al., 1999; Kobak, Taylor, Dottl, Greist, Jefferson, Burroughs, Katzelnick, & Mandell, 1997) showed that, as compared with the alternative methods, IVR caused less embarrassment when depression and anxiety disorders were assessed.

The remaining studies all reported satisfaction with IVR, but without comparison with other methods. So, although IVR may actually be an acceptable methodology, this is an issue that needs further exploration.

## CONCLUSIONS

IVR shows promise in a number of health areas but is yet to be fully explored. IVR can supply automatic population-level information services, but estimating the number of unique users will remain difficult. Patient-monitoring studies have been uncontrolled, suggesting that a comparison with alternative methods may be of interest. Automatic reminder calls, consumption monitoring, psychological studies, and psychological experimentation remain restricted to clinical areas. Behavioral intervention studies were mostly unsuccessful, except when human intervention was used in combination with IVR. There remain unexplored areas that may have an impact on IVR projects—particularly, systematic evaluation of voice, multilingual interfaces, touchphone prevalence, survey response rates, sample bias, use by the elderly, and acceptability.

IVR could be an extremely useful tool in population surveys, since it shares many characteristics with CATI, such as automatic skipping and response validation, while also possessing certain advantages, particularly the lack of automatic operation, greater economy, and standardized interviews in multiple languages. Unlike Web and e-mail surveys, frames may be constructed that have good coverage of household populations. It should be especially useful for assessing sensitive behaviors, since it is likely to be seen as a more confidential method by subjects.

We are of the opinion that IVR technology has considerable potential for survey research and health studies. Although there is some way yet to go, we consider IVR to be an exciting innovation, as well as a potentially economical and flexible tool.

## REFERENCES

- ALBISSER, A. M., HARRIS, R. I., SAKKAL, S., PARSON, I. D., & EN CHAO, S. C. (1996). Diabetes intervention in the information age. *Medical Informatics*, *21*, 297-316.
- ALEMAGNO, S. A., COCHRAN, D., FEUCHT, T. E., STEPHENS, R. C., BUTTS, J. M., & WOLFE, S. A. (1996). Assessing substance abuse treatment needs among the homeless: A telephone-based interactive voice response system. *American Journal of Public Health*, *86*, 1626-1628.
- ALEMAGNO, S. [A.], FRANK, S., MOSAVEL, M., & BUTTS, J. (1998). Screening adolescents for health risks using interactive voice response technology: An evaluation. *Computers in Human Services*, *15*, 27-37.
- ALEMI, F., STEPHENS, R., PARRAN, T., LLORENS, S., BHATT, P., GHADIRI, A., & EISENSTEIN, E. (1994). Automated monitoring of outcomes: Application to treatment of drug abuse. *Medical Decision Making*, *14*, 180-187.
- AMERICAN PSYCHOLOGICAL ASSOCIATION (2001). *PsycINFO* [computer file]. Washington, DC: Author.
- AMERICAN STATISTICAL ASSOCIATION & THE INSTITUTE OF MATHEMATICAL STATISTICS (2001). *Current index to statistics extended database* [computer file]. Bethesda, MD: Author.
- AUSTRALIAN BUREAU OF STATISTICS (1998). *Household use of information technology*. Canberra: Author.
- BACHOFEN, M., NAKAGAWA, A., MARKS, I. M., PARK, J.-M., GREIST, J. H., BAER, L., WENZEL, K. W., PARKIN, J. R., & DOTTL, S. L. (1999). Home self-assessment and self-treatment of obsessive-compulsive disorder using a manual and a computer-conducted telephone inter-

- view: Replication of a U.K.–U.S. study. *Journal of Clinical Psychiatry*, **60**, 545-549.
- BAER, L., JACOBS, D. G., CUKOR, P., O'LAUGHLIN, J., COYLE, J. T., & MAGRUDER, K. M. (1995). Automated telephone screening survey for depression. *Journal of the American Medical Association*, **273**, 1943-1944.
- BARATH, A., & CANNELL, C. F. (1976). Effect of interviewer's voice intonation. *Public Opinion Quarterly*, **40**, 370-373.
- BARDONE, A. M., KRAHN, D. D., GOODMAN, B. M., & SEARLES, J. S. (2000). Using interactive voice response technology and timeline follow-back methodology in studying binge eating and drinking behavior: Different answers to different forms of the same question? *Addictive Behaviors*, **25**, 1-11.
- BURKE, A. (1993). Examining the use of a fully-automated interactive voice response tobacco cessation support line. *American Journal of Health Promotion*, **8**, 93-94, 100.
- CHALMERS, T. C., BERRIER, J., SACKS, H. S., LEVIN, H., REITMAN, D., & NAGALINGAM, R. (1987). Meta-analysis of clinical trials as a scientific discipline: II. Replicate variability and comparison of studies that agree and disagree. *Statistics in Medicine*, **6**, 733-744.
- CHEUNG, S., ENTINE, S. M., & KLOTZ, J. H. (1977). Microcomputer voice-response telephone entry for balanced clinical trial randomization. *Journal of Medical Systems*, **1**, 165-169.
- COCHRAN, W. G. (1977). *Sampling techniques* (3rd ed.). New York: Wiley.
- CORKREY, R., & PARKINSON, L. (2002). A comparison of four computer-based telephone interviewing methods: Getting answers to sensitive questions. *Behavior Research Methods, Instruments, & Computers*, **34**, 354-363.
- COUPER, M. P. (2000). Web surveys: A review of issues and approaches. *Public Opinion Quarterly*, **64**, 464-494.
- DINI, E. F., LINKINS, R. W., & CHANEY, M. (1995). Effectiveness of computer-generated telephone messages in increasing clinic visits. *Archives of Pediatrics & Adolescent Medicine*, **149**, 902-905.
- DYCHES, H., ALEMAGNO, S., LLORENS, S. A., & BUTTS, J. M. (1999). Automated telephone-administered substance abuse screening for adults in primary care. *Health Care Management Science*, **2**, 199-204.
- ERSHOFF, D. H., QUINN, V. P., BOYD, N. R., STERN, J., GREGORY, M., & WIRTSCHAFTER, D. (1999). The Kaiser Permanente prenatal smoking-cessation trial: When more isn't better, what is enough? *American Journal of Preventive Medicine*, **17**, 161-168.
- FRANK, A. P., WANDELL, M. G., HEADINGS, M. D., CONANT, M. A., WOODY, G. E., & MICHEL, C. (1997). Anonymous HIV testing using home collection and telemedicine counselling: A multicenter evaluation. *Archives of Internal Medicine*, **157**, 309-315.
- GÉLINAS-CHEBAT, C., & CHEBAT, J.-C. (1996). Voice and advertising: Effects of intonation and intensity of voice on source credibility, attitudes towards the advertised service and the intent to buy. *Perceptual & Motor Skills*, **83**, 243-262.
- GREIST, J. H., MARKS, I. M., BAER, L., PARKIN, J. R., MANZO, P. A., MANTLE, J. M., WENZEL, K. W., SPIERINGS, C. J., KOBAK, K. A., DOTTL, S. L., BAILEY, T. M., & FORMAN, L. (1998). Self-treatment for obsessive compulsive disorder using a manual and a computerized telephone interview: A U.S.–U.K. study. *MD Computing*, **15**, 149-157.
- HARDING, J. P., HAMM, L. R., EHSANULLAH, R. S. B., HEATH, A. T., SORRELLS, S. C., HAW, J., DUKES, G. E., WOLFE, S. G., MANGEL, A. W., & NORTHCUTT, A. R. (1997). Use of a novel electronic data collection system in multicenter studies of irritable bowel syndrome. *Alimentary Pharmacology & Therapeutics*, **11**, 1073-1076.
- HAVEL, R. D., & WRIGHT, M. P. (1997). Automated interviewing for hepatitis B risk assessment and vaccination referral. *American Journal of Preventive Medicine*, **13**, 392-395.
- HAVICE, M. (1989). How response rates compare for human and digitized phone surveys. *Journalism Quarterly*, **66**, 137-142.
- HAVICE, M. (1990a). Measuring nonresponse and refusals to an electronic telephone survey. *Journalism Quarterly*, **67**, 521-530.
- HAVICE, M. (1990b). Touch-tone polling in a university setting. *College & University*, **65**, 227-234.
- HAVICE, M., & BANKS, M. J. (1991). Live and automated telephone surveys: A comparison of human interviewers and an automated technique. *Journal of the Market Research Society*, **33**, 91-102.
- HYMAN, D. J., HERD, J. A., HO, K. S. I., DUNN, J. K., & GREGORY, K. A. (1996). Maintenance of cholesterol reduction using automated telephone calls. *American Journal of Preventive Medicine*, **12**, 129-133.
- HYMAN, D. J., HO, K. S. I., DUNN, J. K., & SIMONS-MORTON, D. (1998). Dietary intervention for cholesterol reduction in public clinic patients. *American Journal of Preventive Medicine*, **15**, 139-145.
- INFORMATION ACCESS CO. (2001). *Health reference center* [computer file]. Foster City, CA: Author.
- KADISON, P., PELLETIER, E. M., MOUNIB, E. L., OPPEDISANO, P., & POTEAT, H. T. (1998). Improved screening for breast cancer associated with a telephone-based risk assessment. *Preventive Medicine*, **27**, 493-501.
- KEMPE, A., DEMPSEY, C., & POOLE, S. R. (1999). Introduction of a recorded health information line into a pediatric practice. *Archives of Pediatrics & Adolescent Medicine*, **153**, 604-610.
- KOBAK, K. A., GREIST, J. H., JEFFERSON, J. W., MUNDT, J. C., & KATZELNICK, D. J. (1999). Computerized assessment of depression and anxiety over the telephone using interactive voice response. *MD Computing*, **16**, 64-68.
- KOBAK, K. A., TAYLOR, L. H., DOTTL, S. L., GREIST, J. H., JEFFERSON, J. W., BURROUGHS, D., KATZELNICK, D. J., & MANDELL, M. (1997). Computerized screening for psychiatric disorders in an outpatient community mental health clinic. *Psychiatric Services*, **48**, 1048-1057.
- KOBAK, K. A., TAYLOR, L. H., DOTTL, S. L., GREIST, J. H., JEFFERSON, J. W., BURROUGHS, D., MANTLE, J. M., KATZELNICK, 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.
- LAVER, J. D. M. (1968). Voice quality and indexical information. *British Journal of Disorders of Communication*, **3**, 43-54.
- LEIRER, V. O., MORROW, D. G., TANKE, E. D., & PARIANTE, G. M. (1991). Elders' nonadherence: Its assessment and medication reminding by voice mail. *Gerontologist*, **31**, 514-520.
- LIEU, T. A., CAPRA, A. M., MAKOL, J., BLACK, S. B., & SHINEFIELD, H. R. (1998). Effectiveness and cost-effectiveness of letters, automated telephone messages, or both for underimmunized children in a health maintenance organization. *Pediatrics* [On line], **101**. Retrieved from <http://www.pediatrics.org/cgi/content/full/101/104/E103>.
- LINKINS, R. W., DINI, E. F., WATSON, G., & PATRIARCA, P. A. (1994). A randomized trial of the effectiveness of computer-generated telephone messages in increasing immunization visits among preschool children. *Archives of Pediatrics & Adolescent Medicine*, **148**, 908-914.
- MAHONEY, D., TENNSTEDT, S., FRIEDMAN, R., & HEEREN, T. (1999). An automated telephone system for monitoring the functional status of community-residing elders. *Gerontologist*, **39**, 229-234.
- MARKS, I. M., BAER, L., GREIST, J. H., PARK, J. M., BACHOFEN, M., NAKAGAWA, A., WENZEL, K. W., PARKIN, J. R., MANZO, P. A., DOTTL, S. L., & MANTLE, J. M. (1998). Home self-assessment of obsessive-compulsive disorder. *British Journal of Psychiatry*, **172**, 406-412.
- MENEGHINI, L. F., ALBISSER, A. M., GOLDBERG, R. B., & MINTZ, D. H. (1998). An electronic case manager for diabetes control. *Diabetes Care*, **21**, 591-596.
- MILLER, N., MARUYAMA, G., BEABER, R. J., & VALONE, K. (1976). Speed of speech and persuasion. *Journal of Personality & Social Psychology*, **34**, 615-624.
- MORROW, D. [G.], CARVER, L. M., LEIRER, V. O., & TANKE, E. D. (2000). Medication schemas and memory for automated telephone messages. *Human Factors*, **42**, 523-540.
- MORROW, D. G., LEIRER, V. O., CARVER, L. M., TANKE, E. D., & McNALLY, A. D. (1999a). Effects of aging, message repetition, and not-taking on memory for health information. *Journals of Gerontology: Psychological Sciences & Social Sciences*, **54B**, P369-P379.
- MORROW, D. [G.], LEIRER, V. O., CARVER, L. M., TANKE, E. D., & McNALLY, A. D. (1999b). Repetition improves older and younger adult memory for automated appointment messages. *Human Factors*, **41**, 194-204.
- MUNDT, J. C., KELLEHER, P. F., PERRINE, M. W., & SEARLES, J. S. (1997). Psychological performance assessment via interactive voice response



- systems. *Behavior Research Methods, Instruments, & Computers*, **29**, 506-518.
- MUNDT, J. C., KOBAK, K. A., TAYLOR, L. V. H., MANTLE, J. M., JEFFERSON, J. W., KATZELNICK, D. J., & GREIST, J. H. (1998). Administration of the Hamilton depression rating using interactive voice response technology. *MD Computing*, **15**, 1-9.
- MUNDT, J. C., PERRINE, M. W., SEARLES, J. S., & WALTER, D. (1995). An application of interactive voice response (IVR) technology to longitudinal studies of daily behavior. *Behavior Research Methods, Instruments, & Computers*, **27**, 351-357.
- MUNDT, J. C., SEARLES, J. S., PERRINE, M. W., & HELZER, J. E. (1995). Cycles of alcohol dependence: Frequency-domain analyses of daily drinking logs for matched alcohol-dependent and nondependent subjects. *Journal of Studies on Alcohol*, **56**, 491-499.
- NATIONAL LIBRARY OF MEDICINE (2001). *MEDLINE* [Computer file]. Bethesda, MD: Author.
- NICHOLLS, W. L. (1988). Computer-assisted telephone interviewing: A general introduction. In R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls, & J. Waksberg (Eds.), *Telephone survey methodology* (pp. 377-385). New York: Wiley.
- OKSENBERG, L., & CANNELL, C. (1988). Effects of interviewer vocal characteristics on nonresponse. In R. M. Groves, P. P. Biemer, L. E., Lyberg, J. T. Massey, W. L. Nicholls, & J. Waksberg (Eds.), *Telephone survey methodology* (pp. 257-269). New York: Wiley.
- OKSENBERG, L., COLEMAN, L., & CANNELL, C. F. (1986). Interviewers' voices and refusal rates in telephone surveys. *Public Opinion Quarterly*, **50**, 97-111.
- OSGOOD-HYNES, D. J., GREIST, J. H., MARKS, I. M., BAER, L., HENEMAN, S. W., WENZEL, K. W., MANZO, P. A., PARKIN, J. R., SPIERINGS, C. J., DOTTL, S. L., & VITSE, H. M. (1998). Self-administered psychotherapy for depression using a telephone-accessed computer system plus booklets: An open U.S.-U.K. study. *Journal of Clinical Psychiatry*, **59**, 358-365.
- PAGE, R. A., & BALLOUN, J. L. (1978). The effect of voice volume on the perception of personality. *Journal of Social Psychology*, **105**, 65-72.
- PATEL, U. H., & BABBS, C. F. (1992). A computer-based, automated, telephonic system to monitor patient progress in the home setting. *Journal of Medical Systems*, **16**, 101-112.
- PERRINE, M. W., MUNDT, J. C., SEARLES, J. S., & LESTER, L. S. (1995). Validation of daily self-reported alcohol consumption using interactive voice response (IVR) technology. *Journal of Studies on Alcohol*, **56**, 487-490.
- PHIPPS, P. A., & TUPEK, A. R. (1991). Assessing measurement errors in a touchtone recognition survey. *Survey Methodology*, **17**, 15-26.
- PIETTE, J. D. (1999). Patient education via automated calls: A study of English and Spanish speakers with diabetes. *American Journal of Preventive Medicine*, **17**, 138-141.
- PIETTE, J. D., & MAH, C. A. (1997). The feasibility of automated voice messaging as an adjunct to diabetes outpatient care. *Diabetes Care*, **20**, 15-21.
- PIETTE, J. D., MCPHEE, S. J., WEINBERGER, M., MAH, C. A., & KRAEMER, F. B. (1999). Use of automated telephone disease management calls in an ethnically diverse sample of low-income patients with diabetes. *Diabetes Care*, **22**, 1302-1309.
- PIETTE, J. D., WEINBERGER, M., & MCPHEE, S. J. (2000). The effect of automated calls with telephone nurse follow-up on patient-centered outcomes of diabetes care: A randomized, controlled trial. *Medical Care*, **38**, 218-230.
- PIETTE, J. D., WEINBERGER, M., MCPHEE, S. J., MAH, C. A., KRAEMER, F. B., & CRAPO, L. M. (2000). Do automated calls with nurse follow-up improve self-care and glycemic control among vulnerable patients with diabetes? *American Journal of Medicine*, **108**, 20-27.
- PLATEK, R., & GRAY, G. B. (1986). On the definitions of response rates. *Survey Methodology*, **12**, 17-27.
- SCHAEFER, D. R., & DILLMAN, D. A. (1998). Development of a standard e-mail methodology: Results of an experiment. *Public Opinion Quarterly*, **62**, 378-397.
- SCHNEIDER, S. J., SCHWARTZ, M. D., & FAST, J. (1995). Computerized, telephone-based health promotion: I. Smoking cessation program. *Computers in Human Behavior*, **11**, 135-148.
- SEARLES, J. S., HELZER, J. E., & WALTER, D. E. (2000). Comparison of drinking patterns measured by daily reports and timeline follow back. *Psychology of Addictive Behaviors*, **14**, 277-286.
- SEARLES, J. S., PERRINE, M. W., MUNDT, J. C., & HELZER, J. E. (1995). Self-report of drinking using touch-tone telephone: Extending the limits of reliable daily contact. *Journal of Studies on Alcohol*, **56**, 375-382.
- SHEEHAN, K. B., & HOY, M. G. (1999). Using e-mail to survey internet users in the United States: Methodology and assessment. *Journal of Computer Mediated Communication* [On line], **4**(3). Retrieved from <http://www.ascusc.org/jcmc>.
- SMITH, S. M., & SHAFFER, D. R. (1991). Celerity and cajolery: Rapid speech may promote or inhibit persuasion through its impact on message elaboration. *Personality & Social Psychology Bulletin*, **17**, 663-669.
- SOBELL, M. B., & SOBELL, L. C. (1992). Timeline follow-back: A technique for assessing self-reported alcohol consumption. In J. B. Allen (Ed.), *Measuring alcohol consumption: Psychosocial and biological methods* (pp. 41-72). Totowa, NJ: Humana.
- SOCIOLOGICAL ABSTRACTS (2001). *SocioFile* [Computer file]. San Diego: Author.
- SPITZER, R. L., WILLIAMS, J. B. W., KROENKE, K., LINZER, M., DEGRUY, F. V., III, HAHN, S. R., BRODY, D., & JOHNSON, J. G. (1994). Utility of a new procedure for diagnosing mental disorders in primary care: The PRIME-MD 1000 study. *Journal of the American Medical Association*, **272**, 1749-1756.
- STEEL, D., VELLA, J., & HARRINGTON, P. (1996). Quality issues in telephone surveys: Coverage, non-response and quota sampling. *Australian Journal of Statistics*, **38**, 15-34.
- STEHRE-GREEN, P. A., DINI, E. F., LINDGREN, M. L., & PATRIARCA, P. A. (1993). Evaluation of telephoned computer-generated reminders to improve immunization coverage at inner-city clinics. *Public Health Reports*, **108**, 426-430.
- TANKE, E. D., & LEIRER, V. O. (1994). Automated telephone reminders in tuberculosis care. *Medical Care*, **32**, 380-389.
- TANKE, E. D., MARTINEZ, C. M., & LEIRER, V. O. (1997). Use of automated reminders for tuberculin skin test return. *American Journal of Preventive Medicine*, **13**, 189-192.
- TEMPLE, W., TOEWS, J., FIDLER, H., LOCKYER, J. M., TAENZER, P., & PARHOOSINGH, E. J. (1998). Concordance in communication between surgeon and patient. *Canadian Journal of Surgery*, **41**, 439-445.
- TOURANGEAU, R., & SMITH, T. W. (1996). Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly*, **60**, 275-304.
- TROUTMAN, W. G., MURRAY, L. L., & NORLANDER, B. (1990). An estimation of the annual incidence of poisoning using automated telephone polling. *Journal of Toxicology: Clinical Toxicology*, **28**, 193-202.

**APPENDIX A**  
**Study Area for IVR Studies, 1989–2000**

Study	Study Area
Havice (1989)	media surveys
Havice (1990a)	survey refusal rates
Havice (1990b)	survey of student accommodation living conditions
Troutman et al. (1990)	poisoning incidence survey
Havice & Banks (1991)	media surveys
Leirer et al. (1991)	medication nonadherence in the elderly
Phipps & Tupek (1991)	establishment survey
Patel & Babbs (1992)	symptom monitoring for cardiac patients
Baer et al. (1993)	assessment of obsessive compulsive disorder
Burke (1993)	tobacco cessation support line
Stehr-Green et al. (1993)	telephone immunization reminders
Alemi et al. (1994)	monitoring drug client treatment after-care
Linkins et al. (1994)	reminders for childhood immunization
Tanke & Leirer (1994)	automatic reminders for tuberculosis clinic attendance
Baer et al. (1995)	assessment of depression risk
Dini et al. (1995)	telephone clinic appointment reminders
Mundt, Perrine, et al. (1995)	longitudinal monitoring of tobacco and alcohol consumption
Mundt, Searles, et al. (1995)	longitudinal monitoring of alcohol consumption
Perrine et al. (1995)	monitoring of alcohol consumption
Schneider et al. (1995)	automatic smoking self-treatment cessation program
Searles et al. (1995)	longitudinal monitoring of alcohol consumption
Albisser et al. (1996)	diabetes monitoring and feedback
Alemagno et al. (1996)	substance abuse survey among the homeless
Hyman et al. (1996)	cholesterol reduction
Frank et al. (1997)	anonymous HIV testing
Harding et al. (1997)	symptoms monitoring for irritable bowel syndrome
Havel & Wright (1997)	hepatitis B risk assessment and referral
Kobak, Taylor, Dottl, Greist, Jefferson, Burroughs, Mantle, et al. (1997)	assessment of psychological conditions
Kobak, Taylor, Dottl, Greist, Jefferson, Burroughs, Katzelnick, & Mandell (1997)	assessment of psychological conditions
Mundt et al. (1997)	psychological performance assessment
Piette & Mah (1997)	symptom monitoring for diabetes
Tanke et al. (1997)	tuberculin skin test return reminders
Alemagno et al. (1998)	screening adolescents for health risks
Greist et al. (1998)	obsessive-compulsive disorder treatment program
Hyman et al. (1998)	dietary cholesterol reduction
Kadison et al. (1998)	assessment of breast cancer risk
Lieu et al. (1998)	telephone and mail immunization reminders
Marks et al. (1998)	obsessive-compulsive disorder treatment
Meneghini et al. (1998)	diabetes management system
Mundt et al. (1998)	depression risk assessment
Osgood-Hynes et al. (1998)	depression self-treatment program
Temple et al. (1998)	communication between surgeon and patient
Bachofen et al. (1999)	obsessive-compulsive disorder self-treatment
Dyches et al. (1999)	substance abuse screening
Ershoff et al. (1999)	smoking cessation intervention
Kempe et al. (1999)	health information line
Kobak et al. (1999)	assessing depression and anxiety
Mahoney et al. (1999)	monitoring community-residing elders
Piette et al. (1999)	monitoring diabetic patients
Piette (1999)	diabetic information line
Bardone et al. (2000)	monitoring binge eating and drinking
Piette, Weinberger, & McPhee (2000)	diabetes care intervention
Piette, Weinberger, McPhee, et al. (2000)	diabetic self-care intervention
Searles et al. (2000)	daily drinking patterns

**APPENDIX B**  
**Vocal Characteristics**

Characteristic	Effect of Voice	Study
Refusal rates	Interviewers with high pitch variation, fast, loud voices, and clear pronunciation have low refusal rates.	Oksenberg et al. (1986), Oksenberg & Cannell (1988)
Self-report rates	Self-report rates can be influenced by interviewers dropping or raising the tone of their voices at the end of questions.	Barath & Cannell (1976)
Interviewer credibility	Loudness increases perception of credibility, but intonation increases perception of credibility only if the subject feels personally involved in the discussion.	Gélinas-Chebat & Chebat (1996)
Interviewer persuasion	A faster speaking pace is perceived as being more persuasive than a slower pace.	Smith & Shaffer (1991), Miller et al. (1976)
Interviewer confidence	A soft speaker may be perceived as less assured than a speaker with a high volume.	Page & Balloun (1978)

**APPENDIX C**  
**Sample Characteristics for IVR Studies, 1989–2000**

Study	Sample Size	Median Age	Mean Age	% Males	Method
Havice (1989)	575	26–30	35	38	RS
Havice (1990a)	424	–	–	–	RS
Havice (1990b)	1,228	–	–	51	RS
Troutman et al. (1990)	6,478	–	–	–	RS
Havice & Banks (1991)	338; 147	–	37; 35	40; 41	RS
Leirer et al. (1991)	16	–	71	45	CS
Phipps & Tupek (1991)	465	–	–	–	RS
Patel & Babbs (1992)	3	–	–	67	CS
Baer et al. (1993)	18	–	38	44	CS
Burke (1993)	5,345 callers	–	–	–	VS
Stehr-Green et al. (1993)	229	–	9 months	46	RS
Alemi et al. (1994)	42	–	38	48	CS
Linkins et al. (1994)	8,002	–	–	58	CS
Tanke & Leirer (1994)	2,008	19	–	54	CS
Baer et al. (1995)	1,812	–	37; 24	46	VS
Dini et al. (1995)	517	–	–	–	CS
Mundt, Perrine, et al. (1995)	51	28	–	100	CS
Mundt, Searles, et al. (1995)	51	28	–	100	CS
Perrine et al. (1995)	31 couples	36	–	50	VS
Schneider et al. (1995)	37; 975	–	–	–	VS
Searles et al. (1995)	51	28	–	100	CS
Albisser et al. (1996)	204	–	–	–	CS
Alemagno et al. (1996)	207	–	–	76	VS
Hyman et al. (1996)	115	–	47	23	CS
Frank et al. (1997)	1,255	–	–	61	VS
Harding et al. (1997)	270; 370	–	–	–	CS
Havel & Wright (1997)	558	–	27	51	VS
Kobak, Taylor, Dottl, Greist, Jefferson, Burroughs, Mantle, et al. (1997)	200	–	41	33	CS
Kobak, Taylor, Dottl, Greist, Jefferson, Burroughs, Katzelnick, & Mandell (1997)	51	–	44	43	CS
Mundt et al. (1997)	10; 7	33	–	100	CS
Piette & Mah (1997)	74	–	–	95	CS
Tanke et al. (1997)	701	–	45	–	CS
Alemagno et al. (1998)	116	–	–	36	CS
Greist et al. (1998)	40	–	35	53	VS
Hyman et al. (1998)	212	57	–	25	CS
Kadison et al. (1998)	343	30–39	–	0	VS
Lieu et al. (1998)	648	–	–	–	CS

## APPENDIX C (Continued)

Study	Sample Size	Median Age	Mean Age	% Males	Method
Marks et al. (1998)	63	–	33	54	VS
Meneghini et al. (1998)	184	–	–	–	CS
Mundt et al. (1998)	367	–	36	35	CS
Osgood-Hynes et al. (1998)	41	–	42	29	VS
Temple et al. (1998)	8; 26	–	–	–	RS
Bachofen et al. (1999)	21	–	31	57	CS
Dyches et al. (1999)	724	–	–	23	CS
Ershoff et al. (1999)	332	–	29	0	CS
Kempe et al. (1999)	449	–	–	–	VS
Kobak et al. (1999)	113; 72	–	44; 42	–	VS
Mahoney et al. (1999)	20	–	79	15	CS
Piette et al. (1999)	252	–	58	69	CS
Piette (1999)	256	–	58	69	CS
Bardone et al. (2000)	38	–	–	–	CS
Piette, Weinberger, & McPhee (2000)	248	–	55	41	RS
Piette, Weinberger, McPhee, et al. (2000)	280	–	57	49	CS
Searles et al. (2000)	33	–	32	100	CS

Note—RS, random sample; VS, volunteer sample; CS, convenience sample.

(Manuscript received June 6, 2001;  
revision accepted for publication February 1, 2002.)