International Psychogeriatrics (2018), 30:1, 95–102 © International Psychogeriatric Association 2017
doi:10.1017/S1041610217001752

Validation of telephone-based behavioral assessments in aging services clients

Yeates Conwell,[1] Adam Simning,[1] Nicole Driffill,[1] Yinglin Xia,[2] Xin Tu,[3] Susan P. Messing[4] and David Oslin[5]

[1]*Department of Psychiatry, University of Rochester School of Medicine and Dentistry (URSMD), Rochester, New York, USA*
[2]*Department of Medicine, University of Illinois College of Medicine at Chicago, Chicago, Illinois, USA*
[3]*Department of Family Medicine and Public Health, University of California San Diego, La Jolla, California, USA*
[4]*Department of Biostatistics, URSMD, Rochester, New York, USA*
[5]*Department of Psychiatry, University of Pennsylvania and the Cpl Michael J Crescenz VA Medical Center, Philadelphia, Pennsylvania, USA*

ABSTRACT

Background: The Behavioral Health Laboratory (BHL), a telephone-based mental health assessment, is a cost-effective approach that can improve mental illness identification and management. The individual BHL instruments, which were originally designed to be administered in-person, have not yet been validated with an in-person BHL assessment. This study therefore aims to characterize the concordance between the BHL data gathered by telephone and in-person interviews.

Methods: A cross-sectional study was conducted with English-speaking aging services network (ASN) clients aged 60 years and older in Monroe County, NY who were randomized to a BHL interview either in-person ($n = 55$) or by telephone ($n = 53$).

Results: There was strong evidence of equivalence between telephone and in-person interviews for depressive disorders, generalized anxiety, panic disorder, drug misuse, psychosis, PTSD, mental illness symptom severity, and five of the six questions assessing suicidality. There was marginal equivalence in PHQ-9 total scores and one of the six questions assessing suicidal ideation, and no evidence of equivalence between interview modalities for assessing cognitive impairment.

Conclusions: With a few exceptions, the BHL gathered nearly equivalent information via telephone as compared to in-person interviews. This suggests that the BHL may be a cost-effective approach appropriate for dissemination in a wide variety of settings including the ASN. Dissemination of the BHL has the potential to strengthen the linkages between primary care, mental healthcare, and social service providers and improve identification and management of those with late-life mental illness.

Key words: suicide ideation, depression, telephone assessment, survey methods

Introduction

The United States' population is aging and is placing the geriatric mental health services delivery system under considerable strain (Bartels and Naslund, 2013). According to the Institute of Medicine, the "burden of mental illness and substance use disorders in older adults in the United States borders on a crisis" (Institute of Medicine, 2012, pp. ix). Among those aged 65 years and older, 14%–20% have a mental health or substance use condition, and older adult men have the highest suicide rates (Institute of Medicine, 2012). In addition to the considerable morbidity associated with late-life mental illness, the cost to society can be substantial, with late-life depression being associated with a 47%–51% increase in medical costs in ambulatory and inpatient settings (Katon *et al.*, 2003). There is a pressing need for innovative approaches that can be implemented broadly and cost-effectively to address late-life mental healthcare needs. Many innovations are occurring in telehealth, which involve the use of telephones, smartphones, and mobile wireless devices helping millions of patients manage acute and chronic conditions while aiming

Correspondence should be addressed to: Yeates Conwell, MD, Professor of Psychiatry, University of Rochester Medical Center, 300 Crittenden Blvd., Rochester, NY 14642, USA. Phone: (585) 275–6739; Fax: (585) 273–1066. Email: yeates_conwell@urmc.rochester.edu. Received 1 May 2017; revision requested 5 Jun 2017; revised version received 3 Aug 2017; accepted 14 Aug 2017. First published online 20 September 2017.

https://doi.org/10.1017/S1041610217001752 Published online by Cambridge University Press

to increase access to and reduce the cost of care (Dorsey and Topol, 2016).

One approach is the Behavioral Health Laboratory (BHL), which is a cost-effective approach that can improve mental illness identification and management (Oslin *et al.*, 2006). The BHL was developed for use in the Department of Veterans Affairs (VA) primary care clinics and often uses structured telephone assessments to facilitate care decisions and to reduce the burden of in-person visits. When a primary care physician (PCP) has mental health concerns about a patient, the PCP can request a BHL assessment. BHL technicians subsequently conduct a telephone (or in-person)-based interview and provide the PCP a report summarizing the test results with recommendations for clinical care. The BHL assesses several domains: demographics, cognition (for those 55 and older), mania, psychosis, panic disorder, generalized anxiety disorder (GAD), post-traumatic stress disorder (PTSD), alcohol misuse, depression, suicidal ideation, and antidepressant use.

As described by Oslin and colleagues (Oslin *et al.*, 2006), the BHL is cost effective and increases identification of alcohol misuse, drug misuse, suicidal ideation, and depression. While telephone assessments can be an efficient use of resources, effective for tracking mental health symptoms longitudinally, and convenient for older adults (Senior *et al.*, 2007; Claassen *et al.*, 2009), the instruments used in the BHL, including the MINI Neuropsychatric Interview (Sheehan *et al.*, 1998), the Patient Health Questionnaire-9 (PHQ-9) (Kroenke *et al.*, 2001), the Blessed Orientation-Memory-Concentration Test (Katzman *et al.*, 1983), and the Paykel Scale for suicide ideation (Paykel *et al.*, 1974), were validated in-person. Their reliability and validity in the BHL telephone-based assessment, to our knowledge, have yet to be confirmed. Some data suggest that telephone administration of the PHQ-9 yields comparable scores to in-person administration (Pinto-Meza *et al.*, 2005; Ryan *et al.*, 2013), but neither study was conducted in an older adult population. Differences between telephone and in-person administration may exist, with one study reporting that individuals completing the GAD-7 via telephone endorsed more anxiety than those completing it in-person (Ryan *et al.*, 2013). Furthermore, given that the BHL elicits sensitive information regarding alcohol and drug use as well as suicidal ideation, the increased rapport possible with a face-to-face interview may facilitate elicitation of data on these sensitive topics. Alternatively, telephone administration may reduce factors such as social desirability (Senior *et al.*, 2007) thus facilitating more valid responses. There also may be differences in assessing cognitive impairment between telephone and in-person modalities as the in-person interview offers an opportunity for a more comprehensive assessment (e.g. can observe an individual's presentation and environment).

A key limitation of the BHL thus far is that its telephone assessments have yet to be validated with in-person assessments for which the individual survey instruments were originally designed. To test the BHL's procedural validity, we therefore conducted a cross-sectional study to examine the concordance between BHL data gathered by telephone and in-person. We conducted this study in the Aging Services Network (ASN). The ASN is a national system of ~30,000 local and state agencies that annually provide human services to over 10 million older adults and their caregivers to help maintain the independence of older adults (services range from providing information and referral to intensive care management) (O'Shaughnessy, 2008). Consistent with previous research on the concordance of telephone and face-to-face interviews for psychiatric syndromes and symptoms (Fenig *et al.*, 1993; Lee *et al.*, 2010), we hypothesize that there will be strong evidence of equivalence between telephone and in-person interview data for the BHL. Strong evidence of equivalence would suggest that endorsement of psychiatric symptoms does not differ when evaluated via telephone and in-person and would help confirm the BHL's procedural validity.

Methods

Aging services clients aged 60 years and older in Monroe County, NY were referred to the study if case managers (CMs) believed the clients to be "depressed, anxious, or confused" during routine interactions. The CMs, at the time of client intake, assessed clients for potential study eligibility with the sole inclusion criterion at the intake interview being whether the client appeared depressed, anxious, or confused. From 4/1/2009 to 9/15/2010, these clients were then referred to research staff who called the clients. Clients were given a brief description of the study and research staff scheduled in-person meetings for those interested in participating. During the in-person meeting, the study was described in further detail and informed consent was obtained. Individuals were excluded if they were younger than 60 years, unable to communicate in English, or did not have capacity to provide informed consent. Subjects were randomly assigned to an in-person interview conducted in

https://doi.org/10.1017/S1041610217001752 Published online by Cambridge University Press

Table 1. Sociodemographic characteristics of the aging services clients who participated in the in-person or phone interview

SOCIODEMOGRAPHICS	TOTAL (N = 108) PERCENT OR MEAN WITH SD		IN-PERSON (N = 55) PERCENT OR MEAN WITH SD		PHONE (N = 53) PERCENT OR MEAN WITH SD		TEST STATISTIC	*p* VALUE[a]
Age	74.2	9.1	73.6	9.2	74.9	9.2	−0.74	0.461
Gender, female	72.2		72.7		71.7		0.01	0.905
Race, white	85.2		83.6		86.8		0.21	0.644
Marital status, married/ partnered	20.4		21.8		18.9		0.14	0.704
Living alone	58.3		58.2		58.5		0.00	0.974

[a]Two-sided *p* values determined by *t*-test if variable is continuous with degrees of freedom equal to 106 or χ^2 test if variable is categorical with degrees of freedom equal to one.

the subject's home ($n = 55$) or to a telephone interview ($n = 53$). The randomization procedure used block randomization and stratified by gender, interviewer, and age. Participants had an average age of 74.2 years, 72.2% were women, 85.2% were white, 20.4% were currently married, and 58.3% lived alone (Table 1).

Trained research personnel performed interviews in a manner consistent with the procedures employed by previous BHL studies (Oslin *et al.*, 2006). Following the BHL assessment, a standard letter summarizing the combined results of the BHL and CM assessments was sent to the subjects, their primary care providers, and CMs. University of Rochester's Research Subjects Review Board approved this study.

Measures

All measures were administered in a uniform manner on the phone and in-person (e.g. using the same language and research personnel). Scores on a battery of assessments were then compared for agreement using the Food and Drug Administration standard for equivalence, 80% to 125% (Food and Drug Administration, 2015).

Sociodemographics

Information on age, gender, race, marital status, and living arrangement (e.g. living alone) was collected.

Mental health

We used instruments specified by the BHL (Oslin *et al.*, 2006). The Mini-International Neuropsychiatric Interview (i.e. MINI) is a short structured diagnostic interview for DSM-IV and ICD-10 psychiatric disorders (Sheehan *et al.*, 1998). Our study included MINI modules that assessed mania, psychosis, substance misuse, PTSD, GAD, and panic disorder. The MINI has been shown to have good reliability and validity when compared to the Structured Clinical Interview for DSM and is less burdensome (Sheehan *et al.*, 1998). To assess overall mental health, we relied on the SF-12 which is a 12-item instrument that includes a mental health subscale, has demonstrated test–rest validity, and correlates strongly with the longer 36-item survey, the SF-36 (Ware *et al.*, 1996). Depression was assessed with the PHQ-9, a nine-item instrument that measures depression severity and has a sensitivity and specificity of 88% for detecting major depression (Kroenke *et al.*, 2001); we used the pattern of responses on the PHQ-9 to determine if major or minor depression DSM-IV criteria were met. Cognitive impairment was evaluated with the Blessed Orientation-Memory-Concentration Test, a six-item test that includes one memory task, two calculation tasks, and three orientation questions and has adequate sensitivity and specificity for detecting cognitive impairment (Woodford and George, 2007). The Paykel five-item scale was used to assess severity of suicidal ideation (Paykel *et al.*, 1974). All participants were also asked, "During the past month, did you take any medications, prescribed or not, for depression, anxiety or nerves?" For participants endorsing medication use, their medications were reported and subsequently grouped by antidepressant or anxiolytic medication classifications.

Statistical analysis

Differences for continuous and categorical variables in sociodemographic characteristics of the aging services clients who participated in the in-person or phone interview were evaluated with *t*-tests and χ^2 tests, respectively. Equivalence of assessments for continuous outcomes were evaluated employing two one-sided *t*-tests (TOSTs) where *p* values were based on an assumption of a lognormal distribution

https://doi.org/10.1017/S1041610217001752 Published online by Cambridge University Press

of the ratios of geometric means, using pooled error terms, with 95% confidence limits established about these ratios. These limits were then compared with the FDA standard for equivalence (0.80 to 1.25) (Food and Drug Administration, 2015) to determine if the confidence interval associated with the telephone-to-in-person ratio of geometric means fell within this interval. p values < 0.05 were deemed evidence of equivalence. Categorical data utilized TOST χ^2 tests with 95% equivalence limits and margins of $\pm$ 0.2. For the depressive disorders category, the equivalence of the trend between methods of administration was determined by examining Cochran–Armitage Trend Test (a p value of 1.000 indicates strong equivalence). To correct for inflated type I error rates due to multiple comparisons or testing, the false discovery rate (FDR) was used to ensure a correct overall type I error. FDR overcomes the limitation of classic methods such as the Bonferroni correction, which are too conservative for adjusting the type I error for a large number of comparisons or testing as in this study (Benjamini and Hochberg, 1995). Thus, for BHL performance in Table 2, we also report FDR results. All analyses were carried out using SAS/STAT software, Version 9.4 of the SAS System (Copyright © 2002- 2012, SAS Institute Inc.) on a Windows 7 platform.

Results

Sociodemographics

Age, gender, race, marital status, and living arrangement did not differ between those interviewed in-person and by telephone (Table 1).

Mental health

Key components of the interview were analyzed for equivalence and results appear in Table 2. With regard to the prevalence of psychiatric syndromes, major depression was highly prevalent (about one in five participants). There was strong equivalence between the in-person and telephone interviews for major and minor depression as well as for diagnoses of GAD, PTSD, and panic disorder and the presence of substance misuse and psychosis (as evaluated by the MINI).

The SF-12 mental health subscale demonstrated strong equivalence with the in-person interviewees having an average score of 44.5, while those interviewed on the telephone had an average score of 43.5. The PHQ-9 mean scores also showed relatively high levels of depression (mean scores were just under the moderate severity threshold) and indicated marginal equivalence between the in-person and telephone interviews. The assessment of cognitive impairment showed higher scores in the telephone interviews (5.2) compared to the in-person (3.1) interviews, which were not equivalent.

With regard to psychotropic medications, there was marginal equivalence across interview methods for antidepressants with 38.2% of the in-person and 34.0% of the telephone interviewees reporting antidepressants use. Anxiolytic medication use, however, was not equivalent across interview methods (in-person: 14.6%; telephone: 26.4%).

Suicidal ideation, as assessed by the Paykel Scale, showed strong equivalence in four of the five categories of past year suicidality. The exception was the death ideation question (i.e. subjects wished they were dead in the past year), which had marginal equivalence and was endorsed by 32.7% of the in-person and 24.5% of the telephone interviewee. Additionally, 13.0% of the in-person and 13.2% of the telephone interviewees endorsed a lifetime history of a suicide attempt, indicating strong equivalence across assessment modalities.

Discussion

The current study, as an examination of BHL's procedural validity, suggests that dissemination and integration of the BHL into a variety of settings (including the ASN) may represent an efficient approach for strengthening linkages between primary care, mental healthcare, and social service providers. For example, having this service available to ASN providers could meaningfully enhance their ability to screen for, assess, and potentially triage mental health issues. Congruent with our hypothesis, with a few exceptions (marginal equivalence in PHQ-9 total scores and one of the six questions assessing suicidal ideation, and no evidence of equivalence between interview modalities for assessing cognitive impairment), the analyses indicate that assessment of key mental health domains such as psychiatric syndromes, overall mental health, and suicidality is largely equivalent whether evaluated by the BHL in-person or on the telephone. Furthermore, these data can be readily summarized in a report to PCPs and are likely to be highly practical in determining the need further referral, examination, and/or treatment and represent another approach to integration of primary and mental healthcare.

Marginal evidence for equivalence was found for the PHQ-9, antidepressant use, and death ideation item. The reasons for the lack of strong equivalence are unclear. Possibilities include the increased ability to seek clarification and inclusion of non-verbal cues in face-to-face interviews (Irvine

https://doi.org/10.1017/S1041610217001752 Published online by Cambridge University Press

Table 2. Comparison of the BHL assessment by in-person and telephone interview modalities

VARIABLES	IN-PERSON (N=55) PERCENT OR MEAN WITH SD		PHONE (N=53) PERCENT OR MEAN WITH SD		TEST STATISTIC	*p* VALUE[a]	FDR-BASED *p* VALUES[b]	EVIDENCE OF EQUIVALENCE
Syndrome								
Depressive disorders					−0.083	1.000[c]	N/A	Strong
Major depression	20.0%		22.6%					
Minor depression	41.8%		37.7%					
Generalized anxiety disorder	20.0%		18.9%		2.48	0.007	0.029	Strong
Drug misuse	0%		1.9%		−9.69	< 0.001	0.003	Strong
Psychosis	0%		3.8%		−6.20	< 0.001	0.006	Strong
Mania	0%		0%		N/A	N/A	N/A	N/A
PTSD	1.8%		7.6%		−3.52	< 0.001	0.009	Strong
Alcohol	1.8%		1.9%		−7.68	< 0.001	0.012	Strong
Current panic disorder	0%		1.9%		−9.69	< 0.001	0.015	Strong
SF-12 Mental Health	44.5	12.1	43.5	12.6	−3.05	0.002	0.026	Strong
PHQ-9 total score[d]	9.8	5.7	9.6	5.7	1.37	0.087	> 0.050	Marginal
Blessed total score[e]	3.1	4.0	5.2	5.2	−0.60	0.726	> 0.050	None
Psych Meds								
Antidepressant	38.2%		34.0%		1.71	0.044	> 0.050	Marginal
Anxiolytic	14.6%		26.4%		−1.06	0.146	> 0.050	None
Suicidality (last year)								
History of attempt	0%		1.9%		−9.69	<0.001	0.018	Strong
Seriously considered	3.6%		3.8%		−5.46	<0.001	0.021	Strong
Thoughts of suicide	16.4%		13.2%		2.47	0.007	0.032	Strong
Wished were dead	32.7%		24.5%		1.36	0.086	> 0.050	Marginal
Life not worth living	34.6%		35.9%		−2.03	0.021	0.035	Strong
Lifetime history of suicide attempt[f]	13.0%		13.2%		−3.03	0.001	0.024	Strong

[a]*p* values were based on ratios determined with geometric means. To evaluate equivalence between groups, these ratios were used to construct a 95% confidence interval, which was then compared with the FDA standard for equivalence (0.80 to 1.25) to determine if the confidence interval associated with the telephone-to-in-person ratio of geometric means fell within this interval.
[b]These *p* values were determined by using the false discovery rate (FDR) to correct for inflated type I error rates due to multiple comparisons or testing.
[c]For the depressive disorders category, the equivalence of the trend between methods of administration was determined by examining the difference between groups than equivalence (a *p* value of 1.000 indicates strong equivalence).
Some variables were missing information for the *p* value calculation: [d] 55, 52; [f] 54, 53.
[e]Data were exponentiated prior to analyses.

et al., 2013), which may be relevant to death ideation, a concept that can be interpreted in different ways (Van Orden *et al.*, 2013). With regard to the PHQ-9, the small difference between means (in-person: 9.8; telephone: 9.6), while enough to result in marginal statistical equivalence, does not represent a difference that is clinically meaningful.

For the Blessed memory test and anxiolytic use there was no evidence of equivalence. A consideration for the Blessed memory test is that functional issues such as hearing loss may have been magnified by the telephone (telephone interviewees had worse scores) and contributed to the discrepancy between the interview modalities. For the prescribed anxiolytic drugs, a category of medications that is often abused and/or diverted (McLarnon *et al.*, 2011), it is possible that there was decreased reporting in-person as people may be less likely to disclose sensitive information in face-to-face interviews (Reddy *et al.*, 2006).

As with any telephone-based approach, the differences in face-to-face interaction (and assessment) and telephone assessment confer limitations upon the BHL. Most prominently, opportunities for gathering data are more limited during telephone-based assessments (e.g. fewer cues to probe for more information, no ability to use visual cues). Additionally, especially in older adults, telephone interviews may be affected by communication difficulties related to hearing loss and/or cognitive decline. Nonetheless, others have found that psychiatric illness can be equivalently evaluated by phone versus in-person (Rohde *et al.*,

https://doi.org/10.1017/S1041610217001752 Published online by Cambridge University Press

1997; Lee *et al.*, 2010) or internet assessments (Hedman *et al.*, 2013).

In primary care VA settings, the BHL has been shown to be a relatively low-cost approach to assess, monitor, and plan treatment for those with mental health needs (Oslin *et al.*, 2006). Likewise, the BHL has considerable potential for being a cost-effective method to foster mental health services integration in a variety of settings. Given that most mental health screening is conducted by a physician or other treating provider, using the BHL in settings such as ASN, home health, rural/underserved areas, and non-VA outpatient clinics could increase identification, referral, and, possibly, indicated treatment of older adults with mental illness and suicidal ideation. To highlight the BHL's potential, we will use the ASN as a case example. Older adults at risk for suicide are unlikely to seek out help from traditional venues (Conwell and Thompson, 2008), and many older adults who seek ASN services experience thoughts of death and suicide (O'Riley *et al.*, 2014), in addition to having high levels of anxiety and depression (Simning *et al.*, 2010). This is especially pertinent since factors that may cause an older adult to seek help from a community-based provider such as an ASN agency (e.g. functional impairment, social isolation) are associated with an increased risk of completed suicide (Conwell *et al.*, 2000; Fassberg *et al.*, 2012). Yet few ASN clients have seen a mental health specialist prior to the initial ASN intake (Simning *et al.*, 2010). The ASN therefore represents a promising point of contact to bridge the aging services, primary care providers, and mental healthcare systems.

Our study has several limitations that should be considered when evaluating the validity of the BHL. First, due to the sample size, we did not assess how the BHL performed in different sociodemographic subgroups. Second, this study occurred in one ASN agency and only included older adults fluent in English who had sufficient cognitive capacity to provide informed consent, which imposes some limitations on its generalizability. For instance, there are possible differences in access to healthcare or community resources that may affect the operationalization and utility of this model in other regions. Third, we did not exclude participants based on their Blessed score (as was done in the initial testing of the BHL (Oslin *et al.*, 2006)); not excluding these participants is likely to have a minimal impact, however, as our study only included an additional four participants that would have been excluded by the original BHL protocol. Fourth, the BHL was conducted by research staff highly trained in administering mental health measures, whereas BHL technicians who would carry out a "real world" assessment may require additional training to ensure the validity of information gathered via telephone. Fifth, we did not perform formal inter-rater reliability assessments of our two research coordinators, both of whom conducted in-person and telephone-based assessments, which could affect our findings and subsequent interpretations. Lastly, this study was designed to compare interview modalities, and prevalence estimates should be interpreted cautiously because their generalizability is unclear as the aging service's care managers did not record which clients were asked to participate in our study.

Future research is needed to further elucidate the utility and limitations of the BHL and its potential for use in various settings. Because it represents a cost-effective model for mental health issue recognition, referral, decision support, and ongoing monitoring, further research should focus on its effectiveness in different populations and associated outcomes (e.g. reducing hospital readmissions and use of acute services, facilitating treatment adherence) as well as its possible application in the rapidly developing field of telehealth (e.g. increasing access to mental health services in areas that are otherwise limited by scarce resources).

Conclusions and implications

In many countries the population is aging rapidly and innovative models addressing mental health in community-residing older adults is an important component to the improvement of public health. Building upon prior research, our study's findings show that the BHL is able to gather nearly equivalent information via telephone as compared to in-person interviews (although some caution should be exercised with possible over-reporting of cognitive impairment and under-reporting of wishing for one's own death when assessed by telephone). These findings suggest that the BHL may be a cost-effective, innovative model appropriate for dissemination to more fully integrate mental health services in a wide variety of settings. Additional research is warranted, however. In particular, examining the cost-effectiveness of disseminating the BHL into other settings as well as its validity and reliability for tracking changes in mental health outcomes over time would be of high utility. Dissemination of the BHL could help prevent older adults with mental illness who may, because of barriers, stigma, and immobility, otherwise "fall through the cracks" in our current systems of care.

https://doi.org/10.1017/S1041610217001752 Published online by Cambridge University Press

Conflicts of interest

None.

Description of authors' roles

Y. Conwell designed the study, supervised data collection, and helped write the article. A. Simning, N. Driffill, and D. Oslin assisted with the research questions, interpreting the results, and writing the article. Y. Xia, X. Tu, and S.P. Messing assisted with the statistical design of the study and its data analyses.

Acknowledgments

This work was supported in part by National Institute of Mental Health (R24 MH071604) and the VISN 4 Mental Illness Research Education and Clinical Center. Dr. Simning is supported through the Empire Clinical Research Investigator Program, sponsored by the New York State Department of Health.

References

Bartels, S. J. and Naslund, J. A. (2013). The underside of the silver tsunami-older adults and mental health care. *The New England Journal of Medicine*, 368, 493–496.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57, 289–300.

Claassen, C. *et al.* (2009). Telephone-based assessments to minimize missing data in longitudinal depression trials: a project IMPACTS study report. *Contemporary clinical trials*, 30, 13–19.

Conwell, Y. and Thompson, C. (2008). Suicidal behavior in elders. *The Psychiatric Clinics of North America*, 31, 333–356.

Conwell, Y. *et al.* (2000). Completed suicide among older patients in primary care practices: a controlled study. *Journal of the American Geriatrics Society*, 48, 23–29.

Dorsey, E. R. and Topol, E. J. (2016). State of telehealth. *The New England Journal of Medicine*, 375, 154–161.

Fassberg, M. M. *et al.* (2012). A systematic review of social factors and suicidal behavior in older adulthood. *International Journal of Environmental Research and Public Health*, 9, 722–745.

Fenig, S., Levav, I., Kohn, R. and Yelin, N. (1993). Telephone vs. face-to-face interviewing in a community psychiatric survey. *American Journal of Public Health*, 83, 896–898.

Food and Drug Administration (2015). *Approved Drug Products with Therapeutic Equivalence Evaluations*, 35 edn. Silver Spring, MD: United States Food and Drug Administration.

Hedman, E. *et al.* (2013). Telephone versus internet administration of self-report measures of social anxiety, depressive symptoms, and insomnia: psychometric evaluation of a method to reduce the impact of missing data. *Journal of Medical Internet Research*, 15, e229.

Institute of Medicine (2012). *The Mental Health and Substance use Workforce for Older Adults: In Whose Hands?* Washington, DC: The National Academies Press.

Irvine, A., Drew, P. and Sainsbury, R. (2013). 'Am I not answering your questions properly?' clarification, adequacy and responsiveness in semi-structured telephone and face-to-face interviews. *Qualitative Research*, 13, 87–106.

Katon, W. J., Lin, E., Russo, J. and Unutzer, J. (2003). Increased medical costs of a population-based sample of depressed elderly patients. *Archives of General Psychiatry*, 60, 897–903.

Katzman, R., Brown, T., Fuld, P., Peck, A., Schechter, R. and Schimmel, H. (1983). Validation of a short orientation-memory-concentration test of cognitive impairment. *The American Journal of Psychiatry*, 140, 734–739.

Kroenke, K., Spitzer, R. L. and Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16, 606–613.

Lee, S., Tsang, A., Mak, A., Lee, A., Lau, L. and Ng, K. L. (2010). Concordance between telephone survey classification and face-to-face interview diagnosis of one-year major depressive episode in Hong Kong. *Journal of Affective Disorders*, 126, 155–160.

Mclarnon, M. E., Monaghan, T. L., Stewart, S. H. and Barrett, S. P. (2011). Drug misuse and diversion in adults prescribed anxiolytics and sedatives. *Pharmacotherapy*, 31, 262–272.

O'Riley, A. A., Van Orden, K. A., He, H., Richardson, T. M., Podgorski, C. and Conwell, Y. (2014). Suicide and death ideation in older adults obtaining aging services. *The American Journal of Geriatric Psychiatry*, 22, 614–622.

O'Shaughnessy, C. V. (2008). *The Aging Services Network: Accomplishments and Challenges in Serving a Growing Elderly Population*. Washington, DC: National Health Policy Forum.

Oslin, D. W., Ross, J., Sayers, S., Murphy, J., Kane, V. and Katz, I. R. (2006). Screening, assessment, and management of depression in va primary care clinics. The Behavioral Health Laboratory. *Journal of General Internal Medicine*, 21, 46–50.

Paykel, E. S., Myers, J. K., Lindenthal, J. J. and Tanner, J. (1974). Suicidal feelings in the general population: a prevalence study. *British Journal of Psychiatry*, 124, 460–469.

Pinto-Meza, A., Serrano-Blanco, A., Penarrubia, M. T., Blanco, E. and Haro, J. M. (2005). Assessing depression in primary care with the PHQ-9: can it be carried out over the telephone? *Journal of General Internal Medicine*, 20, 738–742.

Reddy, M. K., Fleming, M. T., Howells, N. L., Rabenhorst, M. M., Casselman, R. and Rosenbaum, A. (2006). Effects of method on participants and disclosure rates in research on sensitive topics. *Violence and victims*, 21, 499–506.

Rohde, P., Lewinsohn, P. M. and Seeley, J. R. (1997). Comparability of telephone and face-to-face interviews in

https://doi.org/10.1017/S1041610217001752 Published online by Cambridge University Press

assessing Axis I and II disorders. *The American Journal of Psychiatry*, 154, 1593–1598.

Ryan, T. A., Bailey, A., Fearon, P. and King, J. (2013). Factorial invariance of the patient health questionnaire and generalized anxiety disorder questionnaire. *The British Journal of Clinical Psychology*, 52, 438–449.

Senior, A. C., Kunik, M. E., Rhoades, H. M., Novy, D. M., Wilson, N. L. and Stanley, M. A. (2007). Utility of telephone assessments in an older adult population. *Psychology and Aging*, 22, 392–397.

Sheehan, D. V. *et al*. (1998). 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. *The Journal of Clinical Psychiatry*, 59(Suppl. 20), 22–33.

Simning, A., Richardson, T. M., Friedman, B., Boyle, L. L., Podgorski, C. and Conwell, Y. (2010). Mental distress and service utilization among help-seeking, community-dwelling older adults. *International Psychogeriatrics*, 22, 739–749.

Van Orden, K. A., Simning, A., Conwell, Y., Skoog, I. and Waern, M. (2013). Characteristics and comorbid symptoms of older adults reporting death ideation. *The American Journal of Geriatric Psychiatry*, 21, 803–810.

Ware, J. Jr., Kosinski, M. and Keller, S. D. (1996). A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. *Medical Care*, 34, 220–233.

Woodford, H. J. and George, J. (2007). Cognitive assessment in the elderly: a review of clinical methods. *QJM*, 100, 469–484.

https://doi.org/10.1017/S1041610217001752 Published online by Cambridge University Press