Survey Response Rate and Quality in a Mental Health Clinic Population: Results from a Randomized Survey Comparison

Kelly Stolzmann, MS
Mark Meterko, PhD
Christopher J. Miller, PhD
Lindsay Belanger, MPH
Marjorie Nealon Seibert, MBA
Mark S. Bauer, MD

Abstract

Given the limited ability of informatics-based assessment technologies to reach individuals with serious mental health conditions, this study evaluated the feasibility and data quality of mail-out healthcare surveys in this population to assist in measure selection for a multi-site controlled implementation trial. Veterans were randomly selected from those who had received services at a mental health clinic in the Department of Veterans Affairs, and were randomly assigned to one of three questionnaire lengths. Survey length (48–127 items) was not associated with differences in response rate, percent of items missing, or data quality. However, internal consistency reliability was variable among scales and survey lengths. Additional analyses indicate the above measures of survey data quality may differ among respondents who are non-white and younger and have

Address correspondence to Kelly Stolzmann, MS, Center for Healthcare Organization and Implementation Research, VA Boston Healthcare System, 150 South Huntington Avenue, (152M), Boston, MA 02130, USA. .

Christopher J. Miller, PhD, Center for Healthcare Organization and Implementation Research, VA Boston Healthcare System, 150 South Huntington Avenue, (152M), Boston, MA 02130, USA.

Marjorie Nealon Seibert, MBA, Center for Healthcare Organization and Implementation Research, VA Boston Healthcare System, Boston, MA, USA.

Mark S. Bauer, MD, Center for Healthcare Organization and Implementation Research, VA Boston Healthcare System, Boston, MA, USA.

Mark Meterko, PhD, Performance Measurement, VHA Office of Analytics and Business Intelligence (OABI), ENRM Veterans Affairs Medical Center, Bedford, MA, USA.

Mark Meterko, PhD, Department of Health Law, Policy and Management, Boston University School of Public Health and Management, Boston, USA.

Christopher J. Miller, PhD, Department of Psychiatry, Harvard Medical School, Boston, USA.

Mark S. Bauer, MD, Department of Psychiatry, Harvard Medical School, Boston, USA.

Lindsay Belanger, MPH, Oregon Health and Sciences University, Portland, OR, USA.

Journal of Behavioral Health Services & Research, 2018. 521–532. DOI 10.1007/s11414-018-9617-8

psychotic disorders. These results can inform survey protocols to ensure maximal representation of this vulnerable population in health planning and policy assessment.

Background

Patient-reported process and outcome measures are increasingly recognized in healthcare as important factors in the assessment of patient satisfaction, quality of care, and the impact of organizational and policy changes.[1–6] Assessing the health and satisfaction comprehensively within populations and panels of patients is becoming increasingly important in this age of growth patient-centered medical homes and accountable care organizations[7] which are responsible for providing quality care to defined populations.

Patient-reported measures are often obtained using self-report surveys, despite this methodology having several limitations. Survey fatigue and low response rates can be problematic and potentially degrade the quality of survey-based measures.[8–10] Curiously, there is evidence that survey response rates may have declined in more recent years.[11, 12] Data quality, including missing data and response bias related to survey burden, can also be problematic.[13, 14]

Various types of studies exploring the impact specifically of survey length on response rate and data quality in general populations have shown mixed results. In a review to identify mail questionnaire best practices, Edwards et al. found that response rates were higher when short surveys were used, and that surveys containing questions of a sensitive nature were less likely to be returned.[15] Another study found that the longer the survey, the fewer respondents started and completed the questionnaire.[16] However, in a large random sample, Sheth et al. found no significant differences in response rate between a 4-page (23 items) and a 6-page questionnaire (49 items).[17] Mond and colleagues likewise observed no effect of questionnaire length (8 pages versus 14 pages) on response rate to a mailed survey in a community sample of women.[18]

Notably, surveys have often excluded populations of patients with significant mental health conditions out of concern that response rates would be low and data quality poor. In a commentary on the measurement of patient health-related quality of life, Slade notes that routine outcome assessment is currently not undertaken in most mental health services, and the need exists for cost-effective and efficient methods to collect outcome data in this population.[3] While concerns about cognitive abilities and motivation of individuals with mental health conditions may be reasonable, the authors are aware of no studies of the impact of survey characteristics on response patterns or data quality in this population.

Moreover, despite the increasing use of information technology-based methods (e.g., web-based surveys, smartphone apps, patient-generated outcome data from wearable devices) for patient assessment, work to date indicates that use and desirability of such methods are uneven among those with mental health conditions[19, 20] The use of paper-and-pencil surveys is a logical alternative, although it is concerning that the rates of response to such surveys have been declining over the past 20 years.[12] Continued attention to paper-and-pencil methodologies for those with mental health conditions will be essential in order to ensure adequate representation of individuals treated for serious mental health conditions.

Therefore, the authors conducted a mail-out survey of individuals receiving care for mental health conditions to characterize response rates and patterns, and in particular to identify factors that might compromise the validity and reliability of data from mail surveys in this vulnerable population. Participants were randomly assigned to receive a short, medium, or long survey; the authors compared these groups with regard to completion rates, data quality, and scale psychometrics including Cronbach alpha coefficients and intercorrelation of individual scales.

Methods

Sample

The study sample consisted of veterans receiving ongoing care at a general mental health outpatient clinic in a large Veterans Affairs (VA) Medical Center. The eligible population consisted of all ($n = 1852$) individuals with at least three mental health outpatient visits during a 1-year period (January 2013–2014). Qualifying visits included medication management, individual psychotherapy, and group psychotherapy.

A random sample of 300 veterans was chosen to participate in the study. In order to derive conclusions relevant to non-VA populations, women were oversampled and constituted 45% of those selected for the study, compared to 7% of the eligible VA population. The selected veterans were randomly assigned to one of three survey lengths: lean ($k = 48$ questions), medium ($k = 85$), and maximum ($k = 127$).

Each participant was characterized by up to seven select mental health conditions, based on the presence of at least one ICD-9 code in the medical record during the target period. A count of the total number of diagnoses was computed for each subject, as was occurrence of a mental health hospitalization during the prior year.

Procedures

The survey packet consisted of a cover letter with instructions, the survey, and a pre-paid envelope for the returned questionnaire. In addition, the first mailing included an opt-out postcard, and those who declined participation received no further mailings. Non-respondents received up to two follow-up contacts, mailed at 2-week intervals. No incentives for participation were provided. All study procedures were approved by the VA Central Institutional Review Board.

Table 1
Survey battery components

Measure	No. of pages	No. of items	Lean battery	Medium battery	Maximum battery
Demographics (age, sex, race)	1	3	x	x	x
VR-12 (MCS and PCS)	3	14	x	x	x
Patient Assessment of Chronic Illness Care (PACIC)	2	11	x	x	x
Use of Technology	2	20	x	x	x
VA Mental Health Services Recovery-Oriented Care	1	3		x	x
Cooperative Study #430 Adherence Self-Report	1	1		x	x
AUDIT-C and drug and cigarette use	2	5		x	x
Patient Satisfaction Index	2	12		x	x
Sheehan Disability	2	6		x	x
Health Care Climate	2	10		x	x
BASIS-24	3	24			x
Recovery Assessment Scale	5	18			x
Total	26	127	48	85	127

Measures

The survey consisted of standardized measures of healthcare quality, satisfaction, and outcomes in addition to demographic questions (Table 1). This work was preparatory to a multi-site controlled trial to test facilitation as a means to support the establishment of mental health outpatient teams[21]. In particular, candidate measures were selected in order to assess the domains of satisfaction with care, perceptions of quality of care, mental and physical health status, and quality of life.

Veterans RAND SF-12-item (VR-12)

The VR-12 was developed as a veteran-specific version of the SF-12 measure of overall health-related quality of life[1], [22–24] and yields two validated subscales: the Mental Component Score (MCS) and the Physical Component Score (PCS). Results from a previous study[25] allowed imputation of MCS and PCS scores in the context of missing VR-12 items for select patterns of missing items.

Patient Assessment of Chronic Illness Care short form

The extent to which patients experienced a collaborative approach to care was assessed using the 11-item short form of the Patient Assessment of Chronic Illness Care (PACIC) scale.[26], [27] PACIC is based on the Chronic Care Model[28] and measures the patients' perceptions of the extent to which their providers engage and collaborate with them in problem-solving, goal-setting, and the organization of their care.[29]

Use of Technology Questionnaire

The Use of Technology Questionnaire (UTQ) is a custom measure designed to assess access to technology and interest in using that technology for a variety of health-related tasks, which could be relevant to optimizing future survey methodologies.[19] It queries separately about the respondent's access to three different types of devices (cell phones, tablets, and other computers). It additionally assesses whether the respondent is interested in using each type of device for each of nine different healthcare-related tasks in a dichotomous format. These include tasks related to medication, appointments and laboratories, symptom-related communication, and video-based services.

VA Mental Health Services Recovery-Oriented Care and CSP #430 Adherence Item

The three recovery-oriented care items in the survey reflect how respondents perceive their recovery-oriented care in the VA on a 4-point Likert scale from strongly disagree to strongly agree. The single adherence item asks about taking medications as described by their provider.[30] All items are from a VA healthcare operations quality improvement initiative.

Alcohol Use Disorders Identification Test for Consumption

The Alcohol Use Disorders Identification Test for Consumption (AUDIT-C) is a three-question measure of alcohol consumption designed with the intention of screening for active alcohol use disorders in a primary care setting.[31], [32] Although specific cutoff scores can be used as an indication of heavy drinking, in general, the higher the aggregate score, the greater the likelihood of hazardous drinking habits. In addition to the AUDIT-C, the authors included two additional questions on drug and tobacco use in the survey.[33]

Patient Satisfaction Index

Patient satisfaction was measured via a self-report tool specifically adapted and tested in the mental health setting.[34] The Patient Satisfaction Index was originally constructed from the Satisfaction Index developed and widely tested for use in primary care by Hall et al.[35], [36] Both versions have demonstrated a single main principal component, and the version adapted for mental healthcare demonstrated high internal consistency reliability, high test-retest reliability, and sensitivity to changes in clinical programming.[34]

Sheehan Disability Scale

The Sheehan Disability Scale (SDS) was developed to assess functional impairment in three inter-related domains: work/school, social, and family life. The respondent rates the extent to which work/school, social life, and family responsibilities are impaired by symptoms on a 10-point visual analog scale.[37]

Health Care Climate Questionnaire

The Health Care Climate Questionnaire (HCCQ) is a 10-item measure that evaluates patients' perceptions of their healthcare providers' support for the self-management of mental health conditions.[38], [39] The HCCQ has been validated to assess perception of provider support for self-management, developed for use with individuals with bipolar disorder.[40] For the present study, the measurement tool was modified slightly by substituting the more generic term "mental health condition" for the original "bipolar disorder."

Behavior and Symptom Identification Scale

The Behavior and Symptom Identification Scale (BASIS-24) was developed on psychiatric inpatient, residential, and outpatient populations.[41], [42] It is a 24-item measure permitting assessment of self-reported difficulty in symptoms and functioning. BASIS-24 asks for the degree or frequency of difficulty that the respondent has been experiencing during the past week. The 24 items assess six major areas of difficulty and/or distress: depression/functioning, relationships, self-harm, emotional lability, psychosis, and substance abuse.

Recovery Assessment Scale

The 14 items in the Recovery Assessment Scale assess aspects of resiliency and support. The RAS contains five sections related to Veteran Recovery Assessment, Support for Recovery, Housing, Employment, and Involvement in the Community. All items in the current survey are adapted from a VA healthcare OMHO Veterans Health Outcome Improvement Project.

Statistical analysis

The analytic plan was focused on identifying and characterizing factors affecting survey participation that may be relevant to surveys of among individuals with severe mental health conditions, in order to optimize the validity of future self-report surveys so that future survey efforts can be optimized to validly assess self-reports in this vulnerable population. In this effort, the analyses were guided by Dillman et al.'s total survey error framework and associated goal of maximizing data accuracy within the broader social exchange model of survey response.[43]

The authors first compared the demographic and clinical characteristics of the random sample of 300 survey recipients to the overall study cohort using independent sample *t* tests and chi-square tests. Demographic and clinical characteristics of survey responders and non-responders were similarly compared. Following the definitions established by the American Association for Public Opinion Research,[44] the authors calculated the survey cooperation rate as the percent of completed surveys (COOP1) and the percent of completed and partially completed surveys (COOP2, or adjusted response rate) divided by the total number contacted (i.e., after removing 16 surveys that were undeliverable due to inaccurate address, death, or other reasons for return-to-sender).

The percent of missing items was calculated as the sum of missing survey items divided by the total items in the survey. Survey questions that were not completed as part of a skip pattern and check-all-that-apply-type questions were not included in the percent missing. The percent missing per survey page was similarly computed as the sum of missing survey items on a page divided by the total items on that page for each respondent, excluding check-all-that-apply and questions that were part of a skip pattern. A missing PCS or MCS scale occurred if the pattern of missing items did not allow imputation of scale scores.[25] The Health Care Climate Scale was computed as the sum of the 10 climate items, with at least 7 items needed to receive a score.[40] The AUDIT-C, patient satisfaction, and PACIC scales were computed as the sum of all available scale items.

The percent of missing items and scales by survey length, and differences in mean scale scores across the three survey length groups were compared using analysis of variance (ANOVA). Internal consistency reliability was reported by computing standardized Cronbach alpha coefficients for each scale, and Pearson correlations among scales were also computed to determine whether survey length-related response biases had any effect on the relationships among the measures. All data analysis was generated using SAS software, Version 9.3.

Results

The initial survey mailing resulted in 56 respondents, with an additional 14 after the second mailing and 9 after the third mailing. The overall cooperation rate, including only complete surveys in the numerator (COOP1), was 11% (30/284) and the cooperation rate including both complete and partially completed surveys was 28% (79/284). The eligible target population was aged 53.8 ± 14.8 years, 93% male, and 89% white. The 300 individuals randomly selected to receive a survey were slightly less likely to be white (85 vs 90%; $p = .06$), were younger (50 vs 55 years old; $p < .0001$), and had a lower prevalence of treatment for a psychotic diagnosis (9 vs 13%; $p = .03$) (Table 2).

Compared to the 221 non-respondents, survey responders ($n = 79$) were more likely to be white (95 vs 82%; $p = .02$) and were older (56 vs 48 years old; $p < .0001$), but did not differ on other demographics. Veterans with treatment for psychosis were also somewhat under-represented among responders compared to non-responders (4 vs 10%, $p = .07$), but there were no other significant differences with respect to diagnosis (Table 2). Among the 79 responders, there were no differences in demographic characteristics by survey length.

Response rate did not differ by survey length: lean = 26% (26/100); medium = 25% (25/100); and maximum = 28% (28/100). The average percent of missing items also did not differ (overall ANOVA, $p = .13$). However, the maximum battery had a trend toward more missing items (10.5%) than the lean (4.8%) and medium (5.8%) batteries ($p = .06$). The percent of missing items across the survey did not differ by history of acute or long-term mental health hospitalizations, or by number of diagnoses.

Among the scales present in all three survey lengths, we found no statistically significant differences among the percent missing by survey length (overall ANOVA results: PCS, $p = .41$; MCS, $p = .51$; PACIC, $p = .55$). Mean scores on the PCS, MCS, and PACIC were likewise not significantly different across survey length. Among scales present in both the medium and

Table 2
Comparisons among population subgroups

	Total population ($n = 1852$)	**Subsample mailed survey ($n = 300$)**	**Subsample *not* mailed survey ($n = 1552$)**	**Survey responders ($n = 79$)**	**Survey non-responders ($n = 221$)**
Age (mean ± sd.)	53.8 ± 14.8	50.3 ± 14.1**	54.5 ± 14.9**	55.8 ± 11.5**	48.4 ± 14.4**
Sex (% male)	93%	55%**	100%**	58%	54%
Race (% white)	89%	85%*	90%*	95%**	82%**
Diagnoses (past year; mean ± sd.)	2.0 ± 1.1	2.1 ± 1.1	2.0 ± 1.1	2.1 ± 1.1	2.0 ± 1.1
Anxiety disorder (other than post-traumatic stress disorder)	37%	41%	37%	42%	41%
Post-traumatic stress disorder	58%	56%	59%	58%	56%
Traumatic brain injury	2%	2%	3%	3%	1%
Depression	36%	39%	36%	43%	38%
Bipolar	17%	20%*	16%*	23%	19%
Psychosis	12%	9%**	13%**	4%*	10%*
Substance use disorder	41%	38%	42%	37%	38%
Hospitalization (past year)					
Acute mental health	14%	12%	14%	11%	12%
Long-term mental health	2%	3%	2%	5%	2%
PCS (mean ± sd.)				38.4 ± 11.8	
MCS (mean ± sd.)				36.9 ± 12.5	
PACIC (mean ± sd.)				21.4 ± 7.0	

*$p < .10$
**$p < .05$

maximum batteries, there were no significant differences in the mean scores by survey length; however, the percent missing on the AUDIT-C was 14% among the maximum battery and 0% among the medium battery ($p = .05$). There were no differences by survey length in the percent missing among the Patient Satisfaction Index ($p = .35$) or Health Care Climate Scale ($p = .37$).

Scale internal consistency reliabilities generally exceeded guidelines for adequacy of group comparisons, were comparable to those observed in more general population samples, and were generally consistent across survey lengths (Table 3). However, the MCS and PCS subscales of the VR-12 were exceptions, with observed alpha coefficients lower than previously published

estimates, which were reported as .95 and .96, respectively, for the Veterans SF-36.[45] Kazis et al. also performed Cronbach alpha testing on the four VR-12 subscales, which ranged from .88 to .96.[46] The authors also found differences in Cronbach alpha across survey lengths, with the largest difference in internal consistency reliability observed on the PCS: the medium length survey had an alpha of .73 compared to .31 on the lean survey.

The average percent of missing items by survey page differed among survey lengths (Fig. 1). Respondents did not show the expected pattern of survey fatigue, which would predict a greater percentage of missing items toward the end of the survey. In fact, respondents to the maximum battery had a higher percentage of missing items among the first four pages of the survey compared to latter pages. However, there appeared to be relatively higher rates of missing data on pages 7, 10, and 15. This was associated with small number of items (pages 7 and 15 contained only one and two items, respectively). Page 10 contained items on alcohol use, endorsement of which may have been impacted by social desirability (skipping these items rather than endorsing them). In support of this, on page 10, the AUDIT-C item which asks how often the respondent had six or more drinks on one occasion had the highest percent missing among the maximum survey respondents (39%); the percent missing on this question among the medium length survey respondents was 20%.

Discussion

This study explored response rates and data quality for a mail survey in a population of individuals receiving treatment in a general mental health clinic at a large VA Medical Center. Individuals were randomly assigned to receive one of three surveys of varying length.

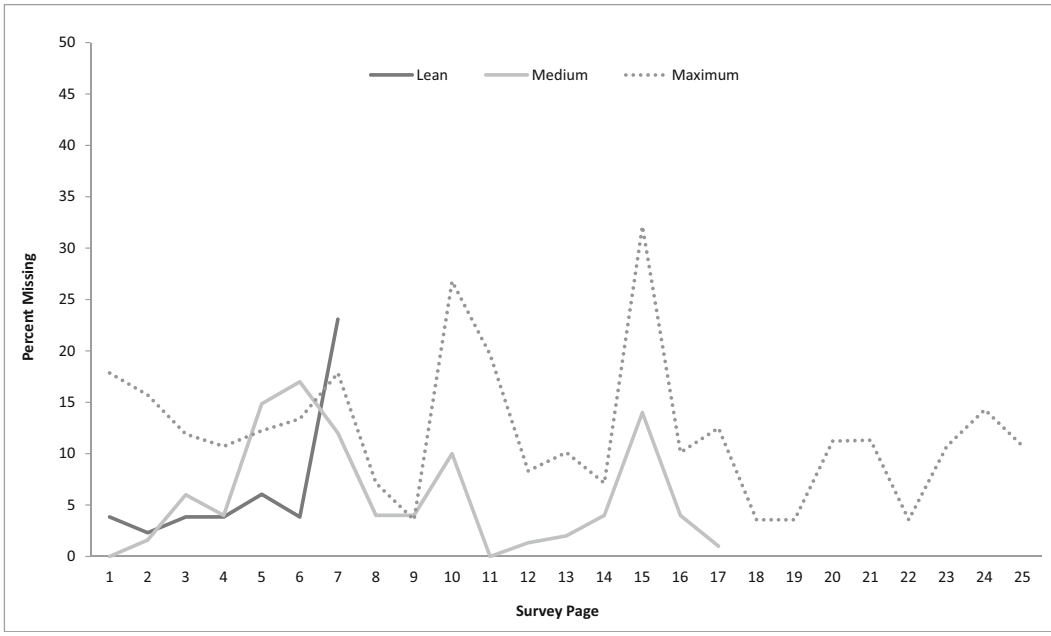
There are four important findings from this study. First, the cooperation rate of 28% was relatively low, but comparable to results from other surveys of similar populations. For example, the Survey of Healthcare Experience of Patients (SHEP) is a patient-based family of assessments of healthcare satisfaction conducted by the VA.[47] The SHEP inpatient survey sample includes veterans discharged from acute care during the previous 6 months. The most recent response rate obtained among the psychiatry inpatients was 20.5%.[48] A recent survey on mental and physical health status following return from deployment to Iraq or Afghanistan received a response from 33% of eligible participants.[49] A response rate of only 16.5% was obtained in a survey mailed to consumers in an outpatient public mental health system.[50] A large national survey among a community sample of elderly, ill patients received a response rate of 25%.[51] Response rates to email surveys have also declined since the late 1980s and analysis of studies that have used both mail and e-mail finds that e-mail does not consistently outperform mailed surveys with respect to survey response rates.[52] In light of the declining response rates among both mail and e-mail survey methods,[12], [52] our response rate may actually be relatively favorable news.

Table 3
Cronbach alphas by survey length

	Lean	Medium	Maximum	Reference
PCS	.31	.73	.33	.96*
MCS	.47	.45	.40	.95*
PACIC	.89	.95	.91	.96
Patient Satisfaction Index		.85	.95	.90
Health Care Climate Questionnaire		.95	.97	.94

*References alphas are for the PCS and MCS for the Veterans SF-36

Figure 1
The average percent of missing items by survey page



It is notable, however, that several characteristics identify subpopulations that may be at risk for under-representation with mail survey methodologies. Specifically, younger individuals (who may prefer electronic methods), non-white individuals, and those with psychotic diagnoses were under-represented in the responder sample compared to the underlying population. Notably, indices of severity, such as the number of diagnoses or recent hospitalization, did not predict non-response, nor did substance use disorders.

Second, the cooperation rate did not differ by survey length. This suggests that the willingness of individuals being treated for mental health conditions to participate in a survey is relatively insensitive to survey length, at least up to 127 items. Thus, researchers and administrators can, with judicious choice, assess several domains of interest using moderate-length instruments in the same survey.

Third, overall, the percent of missing items was quite low, although somewhat higher in the maximum survey. The pattern suggests that there may be some drop-off as survey length increases. However, page-by-page analysis did not indicate that the percent missing simply increased toward the end of longer surveys as might be expected. Rather, the authors observed relatively more missing items earlier in the longer survey. The reason for this is not clear, but may be due to front end skipping to decrease fatigue in anticipation of the overall length of the survey in hand.

Another possible factor to explain the pattern of missing items may be respondent reluctance to respond to items subject to social desirability bias.[14] Among the pages with the highest percent of missing items were questions on page 10, which included the AUDIT-C for characterizing alcohol misuse. In fact, the AUDIT-C item on how often the respondent had six or more drinks on one occasion had the highest percent missing among all survey items. Thus, the pattern of missingness in the current survey may be related to the placement of perceived sensitive items.

Finally, internal consistency reliability of the scales was variable in this population, but generally exceeded guidelines for adequacy of group comparisons. Additionally, coefficient alphas were, for the most part, independent of survey length, indicating that most instruments behaved well independent of overall survey length. However, the exception to this pattern was the VR-12, which demonstrated below-criterion reliability. No immediate explanation for this pattern for this measure is evident. While this may be a product of chance in the context of multiple analyses, the authors returned to the hardcopy survey to explore potential reasons. The authors noted that the physical layout dictated by the VR-12 required numbered items with letter-designated sub-items, as well as changes in anchor characteristics among items (e.g., 1–5 or 1–6 for some, − 2 to + 2 for others). In contrast, the other instruments maintained consistent response anchors throughout. The added complexity of the VR-12 may have taxed the cognitive skills of individuals more burdened with mental health symptoms. The efforts of the NIH PROMIS (Patient-Reported Outcomes Measurement Information System) program may provide a viable alternative for measuring overall physical and mental health status in a more simplified format. This program, an NIH-wide initiative, seeks to develop, test, and disseminate simple condition-specific and cross-condition measures of symptom status, function, and quality of life based on patient experience.[53]

There are several limitations in interpreting these results. First, the data were collected from a single site of a large, publically funded healthcare system, and results may differ in populations treated in other systems. Second, although the authors oversampled women (42% of responders), all respondents were veterans and somewhat more likely to be white, older, and not afflicted with psychotic disorders. Third, the authors purposely utilized a paper-and-pencil mode of administration in view of concerns that individuals treated for mental health conditions may be less likely to respond to web-based surveys; clearly, patient participation and instrument performance may differ with computer-based administration, particularly if item response theory approaches can shorten the instruments without loss of information. Fourth, the sample size of the survey was small, and the power to detect differences by survey length was low. Finally, given the cross-sectional nature of the study, it was not within the scope of the present work to assess the impact of mental health conditions on other aspects of the quality of survey-based data, such as attrition in a longitudinal study.

Overall, though, it is a strength of the study that this sample included a wide variety of individuals with serious mental health conditions, including psychotic disorders, bipolar disorder, and even a small number with traumatic brain injury, with a mean of two active mental health diagnoses. Thus, survey resilience was tested in a more heterogeneous and ill population than one would expect to see in, e.g., a primary care population.

In conclusion, the results from this study are reassuring in demonstrating that individuals in treatment for mental health conditions who respond to surveys using standard clinically relevant self-report instruments can provide good-quality data in terms of relatively little missing data and stable internal consistency reliability for individual scales. However, results also point to certain challenges that administrators and researchers will want to be cognizant of in planning survey assessments of this population. For instance, certain subpopulations may be at risk for under-representation, such as those with psychosis. Additionally, surveys would do best to avoid instruments that utilize a cognitively complex layout, such as the VR-12. The results from this study can serve as the groundwork for future work including the exploration of the impact of mental health conditions on participation, attrition, and data quality in longitudinal studies.

Nonetheless, administrators and researchers can utilize survey approaches in this population with cautious confidence that data quality will be acceptable—although some subpopulations may be under-represented and those scales whose formatting demands the most cognitive effort (e.g., the VR-12) may be vulnerable to variations in performance.

Implications for Behavioral Health

Individuals with severe mental health conditions can be reached by mail-out survey and can provide data comparable, with caveats, to that expected from other groups. This is critical as the authors see the continued growth of patient-centered medical homes[54] and accountable care organizations[7] which are responsible for providing quality care to defined populations. The challenge is to ensure that the voice of such individuals is adequately heard in population- and panel-based assessments of patient-reported outcomes and satisfaction. This methodological report provides guidance to administrators and researchers who work with survey methodology to reach such vulnerable populations.

Compliance with Ethical Standards

All study procedures were approved by the VA Central Institutional Review Board.

Conflict of Interest The authors declare that they have no conflicts of interest.

References

1. Jones D, Kazis L, Lee A, et al. Health status assessments using the Veterans SF-12 and SF-36: methods for evaluating outcomes in the Veterans Health Administration. *The Journal of Ambulatory Care Management.* 2001;24(3):68–86.
2. McCabe R, Saidi M, Priebe S. Patient-reported outcomes in schizophrenia. *The British Journal of Psychiatry Supplement.* 2007;50:s21–28.
3. Slade M. Routine outcome assessment in mental health services. *Psychological Medicine.* 2002;32(8):1339–1343.
4. Cleary PD. The increasing importance of patient surveys. Now that sound methods exist, patient surveys can facilitate improvement. *British Medical Journal.* 1999;319(7212):720–721.
5. Drain M. Quality improvement in primary care and the importance of patient perceptions. *The Journal of Ambulatory Care Management.* 2001;24(2):30–46.
6. Selby JV, Beal AC, Frank L. The Patient-Centered Outcomes Research Institute (PCORI) national priorities for research and initial research agenda. *Journal of the American Medical Association.* 2012;307(15):1583–1584.
7. Shortell SM, Colla CH, Lewis VA, et al. Accountable Care Organizations: The National Landscape. *Journal of Health Politics, Policy and Law.* 2015;40(4):647–668.
8. Fan W, Yan Z. Factors affecting response rates of the web survey: A systematic review. *Computers in Human Behavior.* 2010;26(2):132–139.
9. Groves RM. Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly.* 2006;70(5):646–675.
10. Biemer P, Lyberg L. *Introduction to Survey Quality.* Hoboken, NJ, John Wiley & Sons; 2003.
11. Baruch Y. Response rate in academic studies–a comparative analysis. *Human Relations.* 1999;52(4):421–438.
12. Rookey BD, Le L, Littlejohn M, et al. Understanding the resilience of mail-back survey methods: An analysis of 20 years of change in response rates to national park surveys. *Social Science Research.* 2012;41(6):1404–1414.
13. Rogelberg SG, Stanton JM. Introduction understanding and dealing with organizational survey nonresponse. *Organizational Research Methods.* 2007;10(2):195–209.
14. Tourangeau R, Yan T. Sensitive questions in surveys. *Psychological Bulletin.* 2007;133(5):859–883.
15. Edwards P, Roberts I, Clarke M, et al. Increasing response rates to postal questionnaires: systematic review. *British Medical Journal*. 2002;324:1183.
16. Galesic M, Bosnjak M. Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly.* 2009;73(2):349–360.
17. Sheth JN, Roscoe AM. Impact of questionnaire length, follow-up methods, and geographical location on response rate to a mail survey. *Journal of Applied Psychology.* 1975;60(2):252.
18. Mond JM, Rodgers B, Hay PJ, et al. Mode of delivery, but not questionnaire length, affected response in an epidemiological study of eating-disordered behavior. *Journal of Clinical Epidemiology.* 2004;57(11):1167–1171.
19. Miller CJ, McInnes DK, Stolzmann K, et al. Interest in use of technology for healthcare among Veterans receiving treatment for mental health. *Telemedicine Journal and E- Health*. 2016;22(10):847–854.
20. Shimada SL, Brandt CA, Feng H, et al. Personal health record reach in the Veterans Health Administration: a cross-sectional analysis. *Journal of Medical Internet Research.* 2014;16(12):e272.
21. Bauer MS, Miller C, Kim B, et al. Partnering with health system operations leadership to develop a controlled implementation trial. *Implementation Science.* 2016;11:22.

22. Kazis LE, Ren XS, Lee A, et al. Health status in VA patients: results from the Veterans Health Study. *American Journal of Medical Quality.* 1999;14(1):28–38.
23. McHorney CA, Ware JE, Jr., Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Medical Care.* 1993;31(3):247–263.
24. Ware JE Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Medical Care.* 1992;30(6):473–483.
25. Kazis L, Selim A, Rogers W, et al.. Veterans RAND 12 Item Health Survey (VR-12): A white paper summary. http://www.hosonline.org/surveys/hos/download/Veterans_RAND_12_Item_Health_Survey_White_Paper_Summary.pdf. Accessed April 20, 2015.
26. Vrijhoef HJ, Berbee R, Wagner EH, et al.. Quality of integrated chronic care measured by patient survey: identification, selection and application of most appropriate instruments. *Health Expectations.* 2009;12(4):417–429.
27. Gugiu PC, Coryn C, Clark R, et al. Development and evaluation of the short version of the Patient Assessment of Chronic Illness Care instrument. *Chronic Illness.* 2009;5(4):268–276.
28. Rundall TG, Shortell SM, Wang MC, et al. As good as it gets? Chronic care management in nine leading US physician organisations. *British Medical Journal.* 2002;325(7370):958–961.
29. Schmittdiel J, Mosen DM, Glasgow RE, et al. Patient Assessment of Chronic Illness Care (PACIC) and improved patient-centered outcomes for chronic conditions. *Journal of General Internal Medicine.* 2008;23(1):77–80.
30. Bauer MS, McBride L, Williford WO, et al. Collaborative care for bipolar disorder: part I. Intervention and implementation in a randomized effectiveness trial. *Psychiatric Services.* 2006;57(7):927–936.
31. Bush K, Kivlahan DR, McDonell MB, et al. The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Ambulatory Care Quality Improvement Project (ACQUIP). Alcohol Use Disorders Identification Test. *Archives of Internal Medicine.* 1998;158(16):1789–1795.
32. Bradley KA, Bush KR, Epler AJ, et al. Two brief alcohol-screening tests from the Alcohol Use Disorders Identification Test (AUDIT): validation in a female Veterans Affairs patient population. *Archives of Internal Medicine.* 2003;163(7):821–829.
33. Bauer MS, Krawczyk L, Tuozzo K, et al. Implementing and sustaining team-based telecare for bipolar disorder: lessons learned from a model-guided, mixed methods analysis. *Telemedicine Journal and E-Health.* 2018;24(1):45–53.
34. Nabati L, Shea N, McBride L, et al. Adaptation of a simple patient satisfaction instrument to mental health: psychometric properties. *Psychiatry Research.* 1998;77(1):51–56.
35. Hall JA, Feldstein M, Fretwell MD, et al. Older patients' health status and satisfaction with medical care in an HMO population. *Medical Care.* 1990;28(3):261–270.
36. Hall JA, Milburn MA, Epstein AM. A causal model of health status and satisfaction with medical care. *Medical Care.* 1993;31(1):84–94.
37. Sheehan DV, Harnett-Sheehan K, Raj BA. The measurement of disability. *International Clinical Psychopharmacology.* 1996;11 Suppl 3:89–95.
38. Williams GC, Freedman ZR, Deci EL. Supporting autonomy to motivate patients with diabetes for glucose control. *Diabetes Care.* 1998;21(10):1644–1651.
39. Williams GC, McGregor HA, Sharp D, et al. Testing a self-determination theory intervention for motivating tobacco cessation: supporting autonomy and competence in a clinical trial. *Health Psychology.* 2006;25(1):91–101.
40. Ludman EJ, Simon GE, Rutter CM, et al. A measure for assessing patient perception of provider support for self-management of bipolar disorder. *Bipolar Disorders.* 2002;4(4):249–253.
41. Eisen SV, Normand SL, Belanger AJ, et al. The Revised Behavior and Symptom Identification Scale (BASIS-R): reliability and validity. *Medical Care.* 2004;42(12):1230–1241.
42. Cameron IM, Cunningham L, Crawford JR, et al. Psychometric properties of the BASIS-24(c) (Behaviour and Symptom Identification Scale-Revised) Mental Health Outcome Measure. *International Journal of Psychiatry in Clinical Practice.* 2007;11(1):36–43.
43. Dillman DA Smyth JD, Christian LM. *Internet, Phone, Mail and Mixed-Mode Surveys: The Tailored Design Method (4th edition)*. Hoboken (NJ): John Wiley & Sons; 2014.
44. American Association for Public Opinion Research. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. 8th Edition. 2015; http://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions2015_8theditionwithchanges_April2015_logo.pdf. Accessed May 11,2015.
45. Kazis LE, Lee A, Spiro A, et al. Measurement comparisons of the medical outcomes study and Veterans SF-36 health survey. *Health Care Financing Review.* 2004;25(4):43–58.
46. Kazis LE, Rogers W, Rothendler J, et al. *Outcome Performance Measure Development for Persons with Multiple Chronic Conditions*. Santa Monica, CA. 2017.
47. Cleary PD, Edgman-Levitan S, Roberts M, et al. Patients evaluate their hospital care: a national survey. *Health Affairs.* 1991;10(4):254–267.
48. VHA Office of Analytics and Business Intelligence. Patient Experiences Dimensions of Care, Inpatient Reports. http://vaww.car.rtp.med.va.gov/programs/shep/shepReportsDOCImp.aspx. **Accessed October 5, 2015**.
49. Eisen SV, Schultz MR, Vogt D, et al. Mental and physical health status and alcohol and drug use following return from deployment to Iraq or Afghanistan. *American Journal of Public Health.* 2012;102 Suppl 1:S66–73.
50. Wowra SA, McCarter R. Validation of the empowerment scale with an outpatient mental health population. *Psychiatric Services.* 1999;50(7):959–961.
51. Singer SK, Kerrissey M, Fryer AK, et al. Patients' perspectives on care integration: comparing physician organizations with more and less ACO-like characteristics. Academy of Management; Aug 10, 2015, 2015; Vancouver, Canada.
52. Sheehan K. E-mail survey response rates: a review. *Journal of Computer-Mediated Communication.* 2001;6(2).
53. Hays RD, Bjorner JB, Revicki DA, et al. Development of physical and mental health summary scores from the patient-reported outcomes measurement information system (PROMIS) global items. *Quality of Life Research.* 2009;18(7):873–880.
54. Agency for Healthcare Research and Quality, U.S. Department of Health and Human Services. Coordinating Care for Adults With Complex Care Needs in the Patient-Centered Medical Home: Challenges and Solutions. January 2012.