Brief English and Spanish Survey Detects Change in Response to Advance Care Planning Interventions

Ying Shi, PhD[1,2], **Deborah E. Barnes, PhD, MPH**[2,3,4], **John Boscardin, PhD**[1,4], **John J. You, MD, MSc**[5], **Daren K. Heyland, MD, MSc**[6], **Aiesha M. Volow, MPH**[1,2], **Michelle Howard, PhD, MSc**[7], **Rebecca L. Sudore, MD**[1,2]

[1]Division of Geriatrics, Department of Medicine, University of California, San Francisco, California

[2]San Francisco Veterans Affairs Health Care System, San Francisco, California

[3]Department of Psychiatry, University of California, San Francisco, California

[4]Department of Epidemiology & Biostatistics, University of California, San Francisco, California

[5]Departments of Medicine, and Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada

[6]Department of Critical Care Medicine, Queen's University, Kingston, ON, Canada

[7]Department of Family Medicine, McMaster University, Hamilton, Ontario, Canada

Abstract

Context—The validated 82-item Advance Care Planning (ACP) Engagement Survey measures a broad range of ACP behaviors but is long.

Objectives—Determine whether shorter Survey versions (55-, 34-, 15-, 9-, 4-items) can detect similar change in response to two well-validated ACP interventions and provide practical effect size information.

Methods—We assessed ACP engagement for 986 English- and Spanish-speaking adults in a randomized trial of PREPARE versus an advance directive (AD-only) study arms. The Survey was administered at baseline, 1 week, and 3, 6, 12 months. We calculated mean change scores from baseline to follow-up time points by study arm, intraclass correlation coefficients of change scores between the 82-item Survey with shorter versions, and within- and between-group effect sizes of the mean change scores.

Results—Shorter Survey versions were able to detect within- and between-group changes at all time points. Within-group intraclass correlations of the 82-item to shorter versions were high (0.78 to 0.97) and the amount of between-group differences were comparable using all Survey versions. Twelve-month within-group effect sizes ranged narrowly from 0.76 to 1.05 for different Survey versions in the PREPARE arm and from 0.44 to 0.64 for the AD-only. Between-group effect sizes ranged narrowly from 0.24 to 0.30 for different Survey versions. Results were similar when stratified by English- and Spanish-speakers.

Corresponding author: Ying Shi, PhD, University of California, San Francisco, 4150 Clement Street 151R, San Francisco, CA 94121, Work phone: 415-221-4810 x 23693, Fax: 415-750-6641, Ying.Shi@ucsf.edu.

Conclusion—Shorter versions of the ACP Engagement Survey were able to detect within- and between-group changes comparable to the 82-item version and can be useful for efficiently and effectively measuring ACP engagement in research and clinical settings.

Keywords

Advance care planning; psychometrics; surveys and questionnaires

INTRODUCTION

Advance care planning (ACP) has garnered increasing attention from health systems and researchers because it has been shown to improve patients' satisfaction with medical care and increase agreement of patients' wishes for care received (1–5). Historically, most studies have focused solely on the completion of an advance directive (AD) to measure successful ACP. However, several studies have shown that ACP is a complex process that occurs over time and involves multiple discrete behaviors (6–10). Studies have also shown that people are in varying stages of readiness to engage in these behaviors (11–12).

The ACP Engagement Survey was developed, culturally vetted, and validated to measure the complex process of ACP (13–14). The Survey is based on Social Cognitive and Behavior Change theories and focuses on four behavior change constructs (i.e., knowledge, contemplation, self-efficacy, and readiness) within four ACP domains (i.e., surrogate decision makers, values and quality of life, flexibility in surrogate decision making, and asking doctors questions). Although validated and shown to detect change in response to ACP interventions, the 82-item version of the Survey takes 50 minutes to administer (14–15), reducing its utility. Brief, feasible, validated Surveys that can effectively measure the ACP process and can detect change in response to ACP interventions are needed for research and clinical programs.

In a prior study, we conducted item reduction and validated five progressively shorter versions of the ACP Engagement Survey, including a 55-item, a 34-item, a 15-item, a 9-item, and a 4-item version (14). However, that prior study used blinded trial data with a small sample size and only accessed pre-to-post changes over a 1-week follow-up period.

The current study builds on that prior work by including larger, complete trial cohort data of English- and Spanish-speaking older adults from a published randomized controlled trial designed to compare two well-validated interventions (16). Follow-up time points now include 1 week, 3 months, 6 months, and 12 months, and we calculate both within- and between-group differences by study arm. This study also provides practical effect size information for the use of brief, literacy-appropriate, English and Spanish, culturally vetted measures for a range of ACP behaviors. We will evaluate if change scores in response to an ACP intervention for progressively shorter versions of the Survey, including a 4-item version, are highly correlated with the original 82-item version.

METHODS

Data Sources and Participants

Study participants included 986 English- and Spanish-speaking patients enrolled in a randomized trial at the San Francisco General Hospital from February 2014 through November 2017. These participants were randomly assigned to two intervention groups, an easy-to-read advance directive written at the 5th grade reading level (AD-only arm), and the PREPARE website (PREPAREforYourCare.org) plus the AD (PREPARE arm). PREPARE is an interactive, online ACP program that uses video stories to help people identify their wishes for medical care and models how to discuss those wishes with others. The trial compared the efficacy of PREPARE plus the easy-to-read AD versus the AD alone to engage participants in the ACP process. The study was approved by the University of California, San Francisco Institutional Review Board, and the trial has been published (16–17).

Outcomes and Measures

The validated, patient-reported ACP Engagement 82-item Survey includes 57 items concerning Behavior Change Processes (i.e., knowledge, contemplation, self-efficacy, and readiness) measured on an average five-point Likert scale and 25 ACP Action items such as discussing and documenting ACP wishes using "yes" or "no" response options. The Survey scores were unweighted on a 1 to 5-point scale with higher scores reflecting greater ACP engagement (response options: "1-not at all", "2-a little", "3-somewhat", "4-fairly", and "5-extremely" for knowledge, self-efficacy, and readiness subscales, and "1-never", "2-once or twice", "3-a few times", "4-several times", and "5-a lot" for the contemplation subscale). A detailed table including the questions and response options of the original version and shorter versions (i.e., 55-, 34-, 15-, 9-, and 4-item versions) has been published (14).

The Survey was shortened based on multiple criteria (14). The 25 ACP Action items were removed in all shorter versions due to redundancy because yes/no actions can also be calculated from the readiness questions, which assessed readiness to discuss/document with surrogate decision makers, discuss/document wishes for medical care, discuss/document flexibility for the surrogate, and ask doctors questions. This resulted in all five shorter Survey versions measured on an average five-point Likert scale for the ACP engagement score. The Behavior Change process questions concerning contemplation and questions concerning flexibility in surrogate decision making and asking doctors questions were the questions most often deleted from shorter versions. To be able to compare the average five-point Likert scores of the shorter versions with the original version, the overall average ACP engagement score for the 82-item Survey was created by averaging the five-point Likert scales for the Process measures and also for the Action measures by assigning a value of 5 to response options of "yes" and a value of 0 to response options of "no."

For the randomized trial, we administered the full 82-item ACP Engagement Survey at baseline, 1 week, 3 months, 6 months, and 12 months. We also assessed self-reported participant characteristics at baseline including age, gender, race/ethnicity, education, health literacy, finances and health status (16). In the trial, participants were block-randomized by

adequate versus limited health literacy using a random number generator (18). Prior ACP documentation before the baseline interview was obtained using a composite of any prior legal forms and documented discussions about ACP within the past 5 years by chart review (6,15).

Statistical Analysis

We first compared baseline characteristics of the AD-only and PREPARE arms overall and stratified by English- and Spanish-speakers using unpaired t-tests for continuous variables and Chi-square tests for categorical variables. In addition, we compared baseline characteristics of English- and Spanish-speakers overall. We then assessed the ability of each Survey version to detect change in average ACP Behavior Change Survey scores in response to ACP interventions. We used mixed-effects repeated measures model for the average ACP engagement score with fixed effects on time (i.e., baseline, 1 week, etc.), intervention group (PREPARE versus AD-only arms) and group by time interaction, with time as a categorical variable to allow for non-linearity of responses over time. All analyses were adjusted for the randomization block variable of health literacy and prior ACP documentation and clustered by physician (16). We then calculated mean change scores from baseline to each of the four follow-up time points (i.e., 1 week, and 3, 6, 12 months) by study arm, and measured within-group effect sizes (Cohen's d) of the mean change scores. Using the original 82-item Survey as the reference, we computed the intraclass correlation coefficients of the change in Survey scores at each time point for progressively shorter versions. Finally, we evaluated between-group effect sizes and mean change score differences between PREPARE versus AD-only arms for each shorter Survey version at each follow-up time point and compared them with the original 82-item version using t-tests. All analyses were also stratified by English- and Spanish-speaking participants. We used statistical software SAS 9.4 (SAS Institute Inc) and STATA 15.1 (Stata Corp), all tests of statistical significance were two-sided, and we conducted Bonferroni adjustment for multiple between-group comparisons.

RESULTS

Among 986 enrolled participants, 505 were randomized to the AD-only arm and 481 to the PREPARE arm. The mean age of overall participants was 63.3 (6.4) years, 603 (61.2%) were women, 634 (64.3%) were non-white, 445 (45.1%) were Spanish-speakers, 387 (39.3%) had limited health literacy, 504 (51.1%) reported fair-to-poor health status, and 269 (27.3%) had prior ACP documentation (Table 1). Participant characteristics did not differ between study arms overall or by English- or Spanish-speakers. For the overall cohort including both arms, Spanish-speaking participants were more likely than English-speaking participants to be women, have less education, have higher rates of limited health literacy, and worse self-rated health, p<0.05, Table 1.

Average ACP Behavior Change Survey scores increased more over time in the PREPARE versus the AD-only arms (intervention group by time interaction p<0.001) among all Survey versions (Figure 1), demonstrating that all Surveys can detect change in response to the ACP interventions. The results were similar when stratified by English- and Spanish-speakers (Appendix 1 and 2).

Within-group effect size estimates were larger in the PREPARE versus the AD-only arms at each follow-up time point for all Survey versions (Table 2). Intraclass correlation coefficients of the mean change scores between the original 82-item Survey with progressively shorter versions were medium to high at all follow-up time points for both study arms (range over the four time points 0.78 to 0.97 for PREPARE and 0.76 to 0.98 for AD-only, all p<0.001, Table 2). The 55- and 34-item versions had slightly lower mean change scores compared to the 82-item version, while the 15- and 9-item versions were variable, and the 4-item version had slightly higher values. For example, in the PREPARE arm, the within-group mean change scores with standard deviation (SD) at 12 months compared to baseline were 0.82 (0.9) for the 82-item Survey, 0.73 (0.9) for the 55-item Survey, 0.68 (0.9) for the 34-item Survey, 0.70 (1.0) for the 15-item Survey, 0.75 (1.0) for the 9-item Survey, and 0.91 (1.3) for the 4-item Survey (Table 2). However, these differences between the 82-item version and shorter Surveys were all small and never exceeded 0.23 standard deviation scale (SDs) for PREPARE arm and 0.12 SDs for AD-only arm across all time points (Table 2). Within-group results were similar for English- and Spanish-speaking participants (Appendix 3, 4).

Between-group effect size estimates for PREPARE versus AD-only arms were very similar for all versions of the Survey at all follow-up time points (Table 3, 82-item between-group effect size estimate range over the four time periods 0.24 to 0.31; 55-item 0.21 to 0.26; 34-item 0.21 to 0.24; 15-item 0.20 to 0.25; 9-item, 0.20 to 0.24; 4-item 0.23 to 0.29). As observed for within-group estimates, the 55- and 34-item versions had slightly lower between-group mean change differences compared to the 82-item version while results were mixed with the 15- and 9-item versions, and the 4-item version had slightly higher between-group differences. For example, for PREPARE versus AD-only arms, the between-group differences of mean change scores with SD at 12 months compared to baseline were 0.30 (0.9) for the 82-item Survey, 0.25 (0.9) for the 55-item Survey, 0.25 (0.9) for the 34-item Survey, 0.29 (1.0) for the 15-item Survey, 0.32 (1.0) for the 9-item Survey, and 0.40 (1.2) for the 4-item Survey (Table 3). However, these differences of mean change scores between the 82-item version and shorter Surveys were all small and never exceeded 0.17 SDs across all time points. Results for between-group comparisons of mean change score differences were similar for English- and Spanish-speaking participants (Appendix 5, 6).

DISCUSSION

Using randomized clinical trial data, with four follow-up time points among a large cohort of older adults, we demonstrated that all Survey versions were able to detect change in a broad range of ACP behaviors over time in response to ACP interventions. The Surveys worked well among both English- and Spanish-speaking participants, even though Spanish-speakers had higher rates of limited health literacy and were more likely to have less than a high school education. Having several psychometrically sound shortened versions of the ACP Engagement Survey provides flexibility for research and quality improvement initiatives when choosing Surveys to measure the effectiveness of ACP programs.

We found that the original 82-item version of the ACP Engagement Survey and five progressively shorter versions (i.e., 55-item, 34-item, 15-item, 9-item, and 4-item) can

reliably detect both within- and between-group differences for ACP interventions over all time points (i.e., 1 week, and 3, 6, 12 months). Both within- and between-group effect sizes tended to be higher using the full 82-item Survey, suggesting this version may be most appropriate when maximum power is required, for example for small studies. However, the shorter versions of the Survey were all able to detect both within- and between-group changes, suggesting that they are acceptable alternatives in most clinical and research settings.

This study allowed us to quantify a clinically meaningful change in ACP Engagement Survey scores based on effect sizes using standard thresholds (19). Small effect sizes (0.20– 0.49) were associated with mean change scores of approximately 0.2 to 0.3 points. Moderate effect sizes (0.50–0.79) were associated with mean change scores of approximately 0.4 to 0.5 points. Large effect sizes (0.80) were associated with mean change scores of 0.6 points. Therefore, the smallest clinically meaningful change in response to an ACP intervention would be approximately 0.2 points, and is an evidence that patients are moving along the the behavior change pathway – from pre-contemplation, to contemplation, to preparation, to action. Larger changes of 0.6 or greater likely reflect ACP actions that are farther down the behavior change pathway. For example, in a prior validation study of the Survey in 559 respondents in two countries, a score changes of 1.0 was associated with having completed a prior advance directive (14).

This study also provided detailed within- and between-group effects size information for each version of the Survey at multiple follow-up time points compared to baseline for the overall cohort as well as for English- and Spanish-speakers. These results are important because it will allow ACP researchers to calculate power and estimate sample sizes for future clinical trials. Choice of the Survey version may be based on the length of the Survey desired to reduce response burden, the ACP information important to the research question (as the Behavior Change process questions concerning contemplation, flexibility in surrogate decision making and asking doctors were the most often deleted questions from shorter versions), and the follow-up time proposed (i.e., 1 week, and 3, 6, or 12 months).

The strengths of this study include the rigorous and systematic validation of all Survey versions, assessment of the Survey's ability to detect change over time in response to interventions in English- and Spanish-speakers and use of published trial data. This study does have some limitations. Generalizability may be limited because the validation only took place in one San Francisco health delivery system, with a predominance of older adults. However, the primary care sample was racially and ethnically diverse. Although inclusion criteria required chronic or serious illness for the trial, we do not know whether the results would be similar among patients from specialty clinics or patients who speak a language other than English or Spanish. Future studies will also need to assess the ability of shorter Survey versions to detect change in response to different ACP interventions in varying patient populations and whether the Survey can be used to help tailor ACP discussions based on readiness and behaviors that have not yet been completed in the clinical setting.

In conclusion, progressively shorter versions of the ACP Engagement Survey, including a 4-item version, are psychometrically sound and able to efficiently and effectively measure
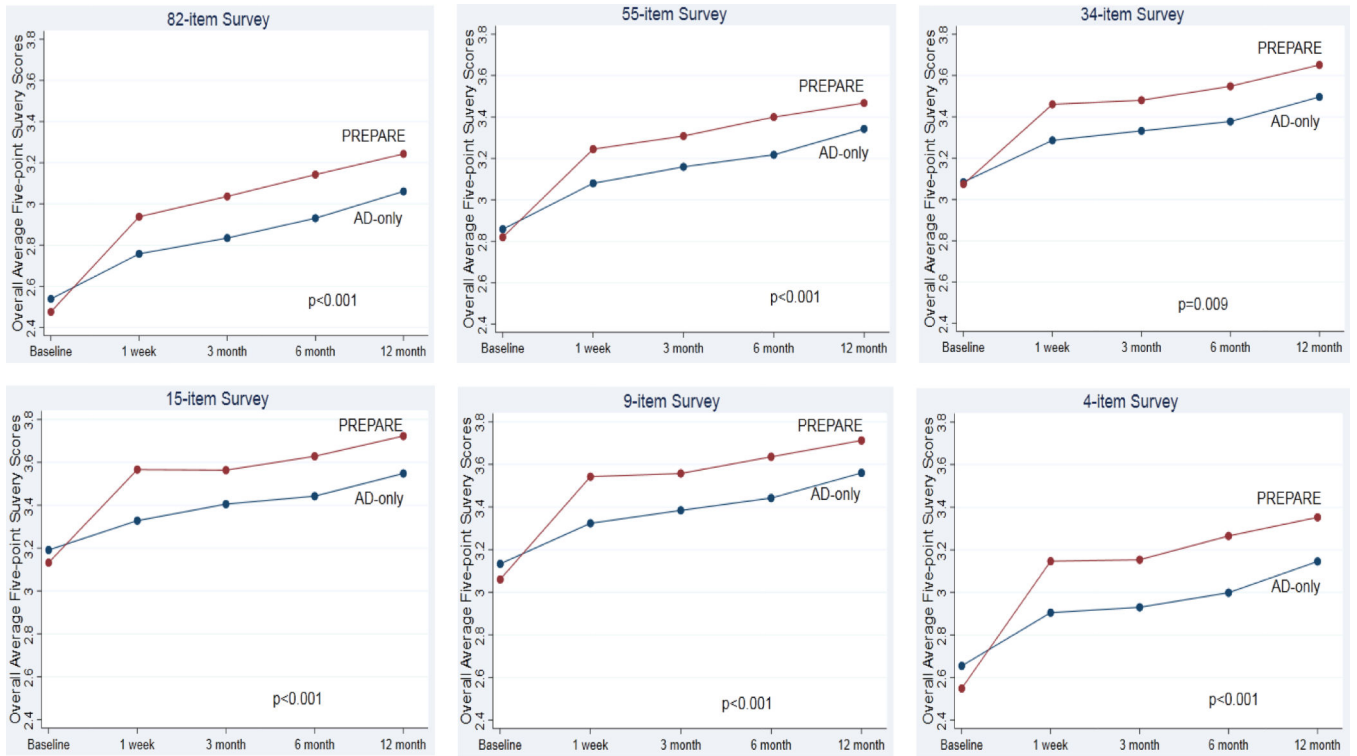
change in ACP behaviors in response to ACP interventions. The choice of which Survey version to use will depend on overall data collection burden, available resources, and the desire to look at Survey subscales or specific Survey domains.

DISCLOSURES AND ACKNOWLEDGMENTS

Appendix

Appendix 1.

English-speaking Participants: Advance Care Planning Engagement Scores at Baseline, 1 Week, 3 Months, 6 Months, and 12 Months for Progressively Shorter Survey Versions by Study Arm[a]

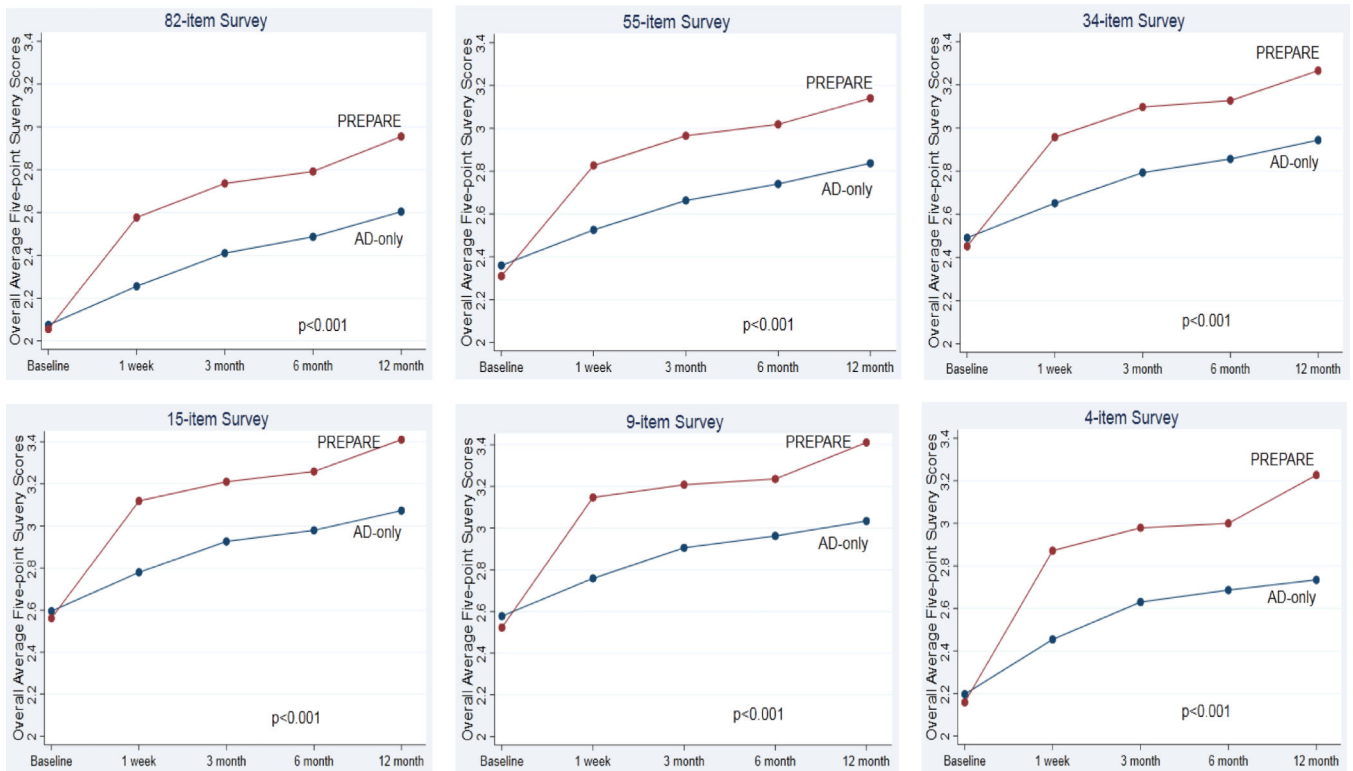[a]P-values in the plots reflect overall intervention group by time interactions.

Appendix 2.

Spanish-speaking Participants: Advance Care Planning Engagement Scores at Baseline, 1 Week, 3 Months, 6 Months and 12 Months for Progressively Shorter Survey Versions by Study Arm[a]

[a]P-values in the plots reflect overall intervention group by time interactions.

Appendix 3.

English-speaking Participants: Within-Group Effect Sizes and Correlation of Mean Change Scores Over Time using Progressively Shorter ACP Engagement Survey Versions

PREPARE (n=262)

	Baseline to 1 Week Follow-up (n=219)[a]			Baseline to 3 Months Follow-up (n=218)[a]			Baseline to 6 Months Follow-up (n=210)[a]			Baseline to 12 Months Follow-up (n=210)[a]		
	Effect Size	Mean Change, Mean (SD)	ICC[b] (95% CI) of Mean Change	Effect Size	Mean Change, Mean (SD)	ICC[b] (95% CI) of Mean Change	Effect Size	Mean Change, Mean (SD)	ICC[b] (95% CI) of Mean Change	Effect Size	Mean Change, Mean (SD)	ICC[b] (95% CI) of Mean Change
82-item	0.57	0.47 (0.7)	------	0.71	0.55 (0.7)	------	0.86	0.66 (0.9)	------	0.95	0.76 (0.8)	------
55-item	0.52	0.44 (0.6)	0.96 (0.95,0.97)	0.62	0.48 (0.7)	0.96 (0.95,0.97)	0.76	0.58 (0.8)	0.97 (0.96,0.98)	0.82	0.65 (0.7)	0.96 (0.95,0.97)
34-item	0.42	0.40 (0.7)	0.94 (0.92,0.95)	0.47	0.40 (0.7)	0.94 (0.92,0.95)	0.56	0.47 (0.8)	0.95 (0.93,0.96)	0.66	0.57 (0.8)	0.94 (0.92,0.95)
15-item	0.45	0.45 (0.8)	0.91 (0.88,0.93)	0.46	0.42 (0.8)	0.90 (0.87,0.93)	0.56	0.49 (0.9)	0.91 (0.89,0.93)	0.64	0.58 (0.9)	0.90 (0.86,0.92)
9-item	0.47	0.49 (0.8)	0.87 (0.83,0.90)	0.50	0.48 (0.8)	0.85 (0.81,0.89)	0.59	0.57 (0.9)	0.89 (0.86,0.92)	0.66	0.65 (0.9)	0.86 (0.82,0.90)
4-item	0.54	0.60 (1.0)	0.78 (0.72,0.83)	0.55	0.58 (1.1)	0.78 (0.72,0.83)	0.67	0.71 (1.2)	0.83 (0.78,0.87)	0.73	0.79 (1.2)	0.80 (0.74,0.85)

AD-only (n=279)

	Baseline to 1 Week Follow-up (n=242)[a]			Baseline to 3 Months Follow-up (n=234)[a]			Baseline to 6 Months Follow-up (n=247)[a]			Baseline to 12 Months Follow-up (n=243)[a]		
	Effect Size	Mean Change, Mean (SD)	ICC[b] (95% CI) of Mean Change	Effect Size	Mean Change, Mean (SD)	ICC[b] (95% CI) of Mean Change	Effect Size	Mean Change, Mean (SD)	ICC[b] (95% CI) of Mean Change	Effect Size	Mean Change, Mean (SD)	ICC[b] (95% CI) of Mean Change
82-item	0.34	0.24 (0.6)	------	0.41	0.33 (0.7)	------	0.49	0.40 (0.8)	------	0.62	0.52 (0.8)	------
55-item	0.34	0.24 (0.6)	0.97 (0.96,0.97)	0.43	0.33 (0.6)	0.97 (0.96,0.97)	0.48	0.37 (0.7)	0.97 (0.96,0.97)	0.61	0.48 (0.8)	0.97 (0.96,0.98)
34-item	0.30	0.22 (0.7)	0.95 (0.93,0.96)	0.34	0.27 (0.7)	0.94 (0.92,0.95)	0.37	0.30 (0.8)	0.94 (0.93,0.96)	0.48	0.41 (0.8)	0.94 (0.93,0.96)
15-item	0.22	0.16 (0.7)	0.91 (0.88,0.93)	0.29	0.25 (0.8)	0.91 (0.88,0.93)	0.30	0.26 (0.8)	0.91 (0.88,0.93)	0.40	0.36 (0.9)	0.92 (0.89,0.94)
9-item	0.25	0.22 (0.8)	0.86 (0.82,0.89)	0.30	0.28 (0.8)	0.87 (0.83,0.90)	0.33	0.32 (0.9)	0.87 (0.84,0.90)	0.42	0.43 (1.0)	0.87 (0.84,0.90)
4-item	0.26	0.27 (1.0)	0.77 (0.70,0.82)	0.27	0.31 (1.0)	0.78 (0.72,0.83)	0.32	0.35 (1.1)	0.82 (0.76,0.86)	0.42	0.50 (1.2)	0.82 (0.77,0.86)

[a] Missing values due to loss-to-follow up at various follow-up time points.

[b] All p-values for intraclass correlation coefficient (ICC) are <0.001.

Appendix 4.

Spanish-speaking Participants: Within-Group Effect Sizes and Correlation of Mean Change Scores Over Time using Progressively Shorter ACP Engagement Survey Versions

PREPARE (n=219)	Baseline to 1 Week Follow-up (n=170)[a]			Baseline to 3 Months Follow-up (n=163)[a]			Baseline to 6 Months Follow-up (n=172)[a]			Baseline to 12 Months Follow-up (n=181)[a]		
	Effect Size	Mean Change, Mean (SD)	ICC[b] (95% CI) of Mean Change	Effect Size	Mean Change, Mean (SD)	ICC[b] (95% CI) of Mean Change	Effect Size	Mean Change, Mean (SD)	ICC[b] (95% CI) of Mean Change	Effect Size	Mean Change, Mean (SD)	ICC[b] (95% CI) of Mean Change
82-item	0.70	0.56 (0.7)	------	0.97	0.69 (0.9)	------	1.03	0.74 (0.9)	------	1.25	0.89 (1.0)	------
55-item	0.69	0.54 (0.7)	0.97 (0.96,0.98)	0.93	0.65 (0.9)	0.98 (0.97,0.98)		0.71 (0.9)	0.98 (0.97,0.98)	1.16	0.83 (1.0)	0.98 (0.97,0.98)
34-item	0.62	0.53 (0.8)	0.95 (0.94,0.97)	0.83	0.64 (0.9)	0.96 (0.95,0.97)	0.86	0.67 (0.9)	0.96 (0.95,0.97)	1.02	0.81 (1.0)	0.97 (0.95,0.97)
15-item	0.63	0.59 (0.8)	0.90 (0.87,0.93)	0.78	0.64 (0.9)	0.92 (0.90,0.94)	0.84	0.68 (1.0)	0.92 (0.90,0.94)	0.99	0.84 (1.0)	0.93 (0.91,0.95)
9-item	0.70	0.65 (0.9)	0.87 (0.82,0.90)	0.80	0.67 (1.0)	0.89 (0.86,0.92)	0.83	0.70 (1.0)	0.91 (0.88,0.93)	1.01	0.88 (1.1)	0.91 (0.89,0.94)
4-item	0.71	0.75 (1.2)	0.77 (0.69,0.83)	0.84	0.81 (1.1)	0.79 (0.71,0.85)	0.85	0.83 (1.3)	0.79 (0.72,0.84)	1.07	1.05 (1.3)	0.82 (0.75,0.86)

AD-only (n=226)	Baseline to 1 Week Follow-up (n=198)[a]			Baseline to 3 Months Follow-up (n=182)[a]			Baseline to 6 Months Follow-up (n=187)[a]			Baseline to 12 Months Follow-up (n=197)[a]		
	Effect Size	Mean Change, Mean (SD)	ICC[b] (95% CI) of Mean Change	Effect Size	Mean Change, Mean (SD)	ICC[b] (95% CI) of Mean Change	Effect Size	Mean Change, Mean (SD)	ICC[b] (95% CI) of Mean Change	Effect Size	Mean Change, Mean (SD)	ICC[b] (95% CI) of Mean Change
82-item	0.24	0.20 (0.6)	------	0.47	0.35 (0.7)	------	0.54	0.41 (0.8)	------	0.72	0.53 (1.0)	------
55-item	0.21	0.18 (0.6)	0.97 (0.96,0.98)	0.41	0.32 (0.7)	0.97 (0.96,0.98)	0.50	0.39 (0.8)	0.98 (0.97,0.98)	0.65	0.48 (1.0)	0.98 (0.97,0.99)
34-item	0.20	0.17 (0.7)	0.94 (0.92,0.96)	0.40	0.31 (0.8)	0.95 (0.94,0.97)	0.45	0.37 (0.8)	0.96 (0.94,0.97)	0.58	0.45 (1.0)	0.98 (0.97,0.98)
15-item	0.21	0.20 (0.8)	0.89 (0.86,0.92)	0.40	0.34 (0.8)	0.92 (0.89,0.94)	0.43	0.39 (0.9)	0.90 (0.86,0.92)	0.55	0.47 (1.0)	0.94 (0.92,0.96)
9-item	0.20	0.19 (0.9)	0.82 (0.77,0.87)	0.39	0.33 (0.9)	0.87 (0.82,0.90)	0.42	0.39 (0.9)	0.87 (0.82,0.90)	0.51	0.45 (1.1)	0.92 (0.90,0.94)
4-item	0.26	0.27 (1.0)	0.74 (0.66,0.80)	0.46	0.44 (1.1)	0.80 (0.74,0.85)	0.49	0.49 (1.0)	0.81 (0.74,0.86)	0.56	0.52 (1.2)	0.86 (0.82,0.90)

[a] Missing values due to loss-to-follow up at various follow-up time points.

[b] All p-values for intraclass correlation coefficient (ICC) are <0.001.

Appendix 5.

English-speaking Participants: Between-Group Effect Sizes and Differences of Mean Change Scores Over Time using Progressively Shorter ACP Engagement Survey Versions

PREPARE versus AD-only	Baseline to 1 Week Follow-up		Baseline to 3 Months Follow-up		Baseline to 6 Months Follow-up		Baseline to 12 Months Follow-up	
	Effect Size of Differences	Differences of Mean Change, Mean (SD)^a	Effect Size of Differences	Differences of Mean Change, Mean (SD)^a	Effect Size of Differences	Differences of Mean Change, Mean (SD)^a	Effect Size of Differences	Differences of Mean Change, Mean (SD)^a
82-item	0.13	0.23 (0.7)	0.20	0.22 (0.7)	0.26	0.26 (0.8)	0.23	0.25 (0.8)
55-item	0.12	0.20 (0.6)	0.13	0.16 (0.7)	0.22	0.21 (0.8)	0.15	0.17 (0.8)
34-item	0.11	0.18 (0.7)	0.12	0.12 (0.7)	0.19	0.16 (0.8)	0.18	0.16 (0.8)
15-item	0.18	0.28 (0.8)	0.12	0.17 (0.8)	0.20	0.23 (0.9)	0.19	0.22 (0.9)
9-item	0.14	0.28 (0.8)	0.12	0.20 (0.8)	0.18	0.25 (0.9)	0.16	0.22 (1.0)
4-item	0.15	0.33 (1.0)	0.16	0.26 (1.0)	0.23	0.36 (1.1)	0.18	0.29 (1.2)

^{a}T-tests for comparing differences of mean change between progressively shorter Survey versions and the original 82-item version all had non-significant p-values with Bonferroni adjustment for multiple comparisons at a significance level of 0.05, which meant no obvious differences among Survey versions.

Appendix 6.

Spanish-speaking Participants: Between-Group Effect Sizes and Differences of Mean Change Scores Over Time using Progressively Shorter ACP Engagement Survey Versions.

PREPARE versus AD-only	Baseline to 1 Week Follow-up		Baseline to 3 Months Follow-up		Baseline to 6 Months Follow-up		Baseline to 12 Months Follow-up	
	Effect Size of Differences	Differences of Mean Change, Mean (SD)^a	Effect Size of Differences	Differences of Mean Change, Mean (SD)^a	Effect Size of Differences	Differences of Mean Change, Mean (SD)^a	Effect Size of Differences	Differences of Mean Change, Mean (SD)^a
82-item	0.37	0.36 (0.7)	0.40	0.33 (0.8)	0.37	0.32 (0.9)	0.37	0.37 (1.0)
55-item	0.35	0.36 (0.7)	0.36	0.33 (0.8)	0.33	0.32 (0.8)	0.36	0.35 (1.0)
34-item	0.34	0.36 (0.7)	0.30	0.32 (0.8)	0.32	0.30 (0.9)	0.35	0.36 (1.0)
15-item	0.39	0.39 (0.8)	0.31	0.30 (0.9)	0.29	0.30 (0.9)	0.38	0.37 (1.0)
9-item	0.39	0.46 (0.9)	0.31	0.34 (0.9)	0.29	0.31 (1.0)	0.38	0.43 (1.1)
4-item	0.40	0.48 (1.1)	0.34	0.38 (1.1)	0.32	0.34 (1.1)	0.47	0.52 (1.2)

^{a}T-tests for comparing differences of mean change between progressively shorter Survey versions and the original 82-item version all had non-significant p-values with Bonferroni adjustment for multiple comparisons at a significance level of 0.05, which meant no obvious differences among Survey versions.

REFERENCES

1. Klingler C, In der Schmitten J, Marckmann G. Does facilitated advance care planning reduce the costs of care near the end of life? Systematic review and ethical considerations. Palliat Med 2016;30:423–433. [PubMed: 26294218]

2. Silveira MJ, Kim SY, Langa KM. Advance directives and outcomes of surrogate decision making before death. N Engl J Med 2010;362:1211–1218. [PubMed: 20357283]

3. Hammes BJ, Rooney BL, Gundrum JD. A comparative, retrospective, observational study of the prevalence, avail-ability, and specificity of advance care plans in a county that implemented an advance care planning microsystem. J Am Geriatr Soc 2010;58:1249–1255. [PubMed: 20649688]

4. Detering KM, Hancock AD, Reade MC, Silvester W. The impact of advance care planning on end of life care in elderly patients: randomised controlled trial. BMJ 2010;340:c1345. [PubMed: 20332506]

5. Pope TM. Legal briefing: Medicare coverage of advance care planning. J Clin Ethics 2015;26(4): 361–367. [PubMed: 26752396]

6. Sudore RL, Fried TR. Redefining the "planning" in advance care planning: preparing for end-of-life decision making. Ann Intern Med 2010;153:256–261. [PubMed: 20713793]

7. Sudore RL, Schickedanz AD, Landefeld CS, et al. Engagement in multiple steps of the advance care planning process: a descriptive study of diverse older adults. J Am Geriatr Soc 2008;56:1006–1013. [PubMed: 18410324]

8. Fried TR, Redding CA, Robbins ML, et al. Stages of change for the component behaviors of advance care planning. J Am Geriatr Soc 2010;58:2329–2336. [PubMed: 21143441]

9. Fried TR, Redding CA, Robbins ML, et al. Promoting advance care planning as health behavior change: development of scales to assess decisional balance, medical and religious beliefs, and processes of change. Patient Educ Couns 2012;86:25–32. [PubMed: 21741194]

10. Sudore RL, Heyland DK, Lum HD, et al. Outcomes that define successful advance care planning: a Delphi panel consensus. J Pain Symptom Manage 2018;55(2):245–255. [PubMed: 28865870]

11. Sudore RL, Knight SJ, McMahan RD, et al. A novel website to prepare diverse older adults for decision making and advance care planning: a pilot study. J Pain Symptom Manage 2014;47(4): 674–86. [PubMed: 23972574]

12. Lum HD, Barnes DE, Katen MT, et al. Improving a full range of advance care planning behavior change and action domains: the PREPARE randomized trial. J Pain Symptom Manage 2018;56(4): 575–581. [PubMed: 29940209]

13. Sudore RL, Stewart AL, Knight SJ, et al. Development and validation of a questionnaire to detect behavior change in multiple advance care planning behaviors. PLoS One 2013;8:e72465. [PubMed: 24039772]

14. Sudore RL, Heyland DK, Barnes DE, et al. Measuring advance care planning: optimizing the advance care planning engagement Survey. J Pain Symptom Manage 2017;53(4):669–681. [PubMed: 28042072]

15. Howard M, Bonham AJ, Heyland DK, et al. Measuring engagement in advance care planning: a cross-sectional multicentre feasibility study. BMJ Open 2016;6(6):e010375.

16. Sudore RL, Schillinger D, Katen MT, et al. Engaging diverse English- and Spanish-speaking older adults in advance care planning: the PREPARE randomized clinical trial. JAMA Intern Med 2018;178(12):1616–1625. [PubMed: 30383086]

17. Sudore RL, Barnes DE, Le GM, et al. Improving advance care planning for English-speaking and Spanish-speaking older adults: study protocol for the PREPARE randomized controlled trial. BMJ Open 2016;6(7): e011705.

18. Chew LD, Griffin JM, Partin MR, et al. Validation of screening questions for limited health literacy in a large VA outpatient population. J Gen Intern Med 2008;23(5):561–566. [PubMed: 18335281]

19. Cohen J Statistical Power Analysis for the Behavioral Sciences. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
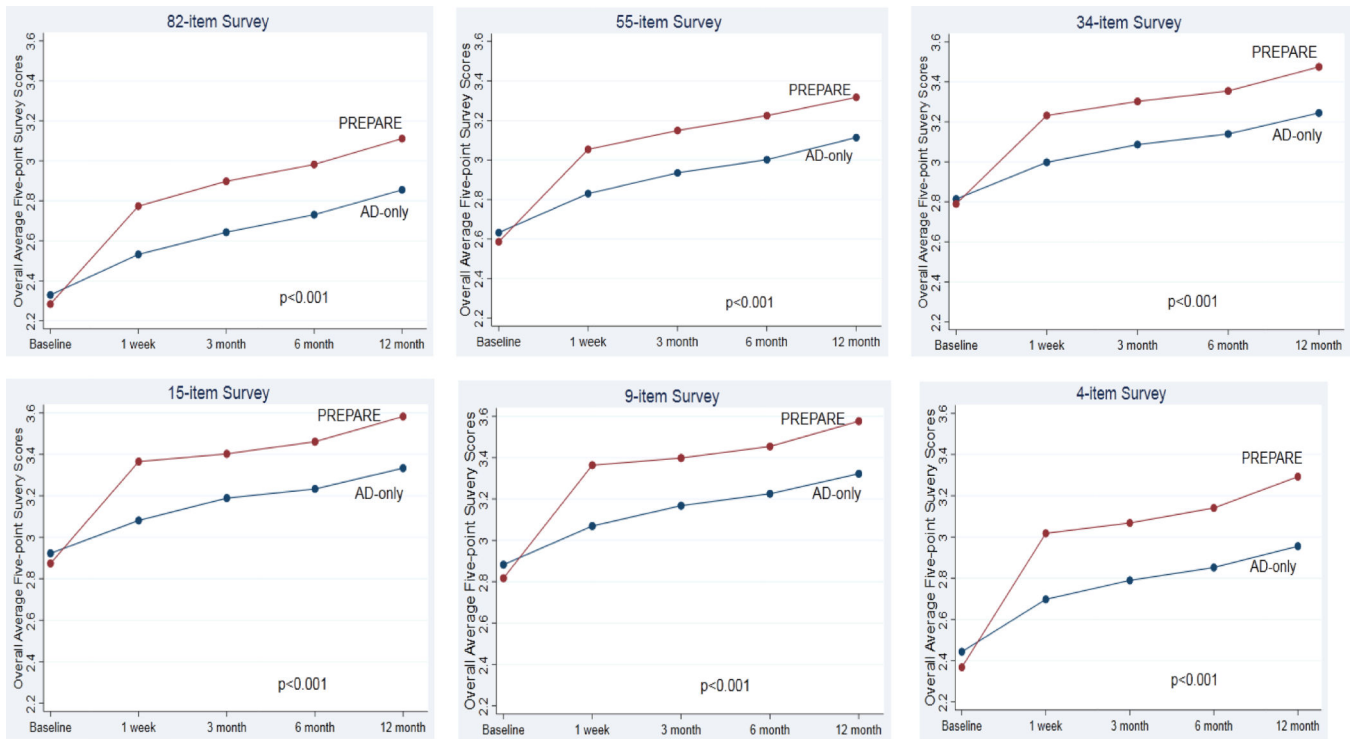
Figure 1.
Progressively Shorter Advance Care Planning Engagement Survey Versions Are Able to Detect Change at 1 Week, 3 Months, 6 Months, and 12 Months by Study Arm[a][b]
[a]English-speakers and Spanish-speakers had similar results as shown in the Appendix 1 and 2.
[b]P-values in the plots reflect overall intervention group by time interactions.

Baseline Participant Characteristics

Table 1.

Participant Characteristic	All Participants (N=986) No. (%)		English Speakers (N=541) No. (%)		Spanish Speakers (N=445) No. (%)	
	AD-only (n=505)	PREPARE (n=481)	AD-only (n=279)	PREPARE (n=262)	AD-only (n=226)	PREPARE (n=219)
Age, mean (SD)	63.1 (6.3)	63.5 (6.4)	62.3 (5.4)	63.0 (6.1)	64.1 (7.2)	64.1 (6.8)
Gender						
Women	314 (62.2)	289 (60.1)	151 (54.1)	132 (50.4)	163 (72.1)	157 (71.7)
Men	191 (37.8)	192 (39.9)	128 (45.9)	130 (49.6)	63 (27.9)	62 (28.3)
Race/Ethnicity:						
White Latino or Hispanic	248 (49.1)	251 (52.2)	24 (8.6)	35 (13.4)	224 (99.1)	216 (98.6)
White non-Latino	104 (20.6)	85 (17.7)	104 (37.3)	84 (32.1)	0	1 (0.5)
African American	92 (18.2)	86 (17.9)	92 (33.0)	86 (32.8)	0	0
Asian/Pacific Islander	34 (6.7)	44 (9.1)	34 (12.2)	44 (16.8)	0	0
Multiethnic/other	27 (5.4)	15 (3.1)	25 (8.9)	13 (4.9)	2 (0.9)	2 (0.9)
Education high school	287 (56.8)	289 (60.1)	102 (36.6)	102 (38.9)	185 (81.9)	187 (85.4)
Limited Health Literacy[a]	202 (40.3)	185 (39.0)	60 (21.7)	56 (21.7)	142 (63.1)	129 (59.7)
Finances[a], not enough to make ends meet	124 (25.0)	119 (25.1)	65 (23.8)	56 (21.5)	59 (26.5)	63 (29.4)
Self-Rated Health[a], fair-to-poor	249 (49.4)	255 (53.2)	122 (43.9)	128 (49.0)	127 (56.2)	127 (58.3)
Prior ACP Documentation	148 (29.3)	121 (25.2)	84 (30.1)	77 (29.4)	64 (28.3)	44 (20.1)

[a] Missing values: limited health literacy: 1.1% missing; finances: 1.6% missing; self-rated health: 0.3% missing.

Table 2.

Within-Group Effect Sizes and Correlation of Mean Change Scores Over Time using Progressively Shorter ACP Engagement Survey Versions [a]

PREPARE (n=481)

	Baseline to 1 Week Follow-up (n=389) [b]			Baseline to 3 Months Follow-up (n=381) [b]			Baseline to 6 Months Follow-up (n=382) [b]			Baseline to 12 Months Follow-up (n=391) [b]		
	Effect Size	Mean Change, Mean (SD)	ICC[c] (95% CI) of Mean Change	Effect Size	Mean Change, Mean (SD)	ICC[c] (95% CI) of Mean Change	Effect Size	Mean Change, Mean (SD)	ICC[c] (95% CI) of Mean Change	Effect Size	Mean Change, Mean (SD)	ICC[c] (95% CI) of Mean Change
82-item	0.62	0.51 (0.7)	------	0.82	0.61 (0.8)	------	0.91	0.70 (0.9)	------	1.05	0.82 (0.9)	------
55-item	0.57	0.48 (0.7)	0.97 (0.96,0.97)	0.73	0.55 (0.8)	0.97 (0.96,0.98)	0.82	0.64 (0.8)	0.97 (0.97,0.98)	0.92	0.73 (0.9)	0.97 (0.96,0.98)
34-item	0.48	0.46 (0.7)	0.95 (0.94,0.96)	0.60	0.50 (0.8)	0.95 (0.94,0.96)	0.66	0.56 (0.9)	0.95 (0.94,0.96)	0.77	0.68 (0.9)	0.95 (0.94,0.96)
15-item	0.53	0.51 (0.8)	0.91 (0.88,0.92)	0.60	0.51 (0.9)	0.91 (0.89,0.93)	0.66	0.58 (0.9)	0.92 (0.90,0.93)	0.76	0.70 (1.0)	0.92 (0.90,0.93)
9-item	0.54	0.56 (0.9)	0.87 (0.84,0.89)	0.61	0.56 (0.9)	0.87 (0.85,0.90)	0.66	0.63 (1.0)	0.90 (0.88,0.92)	0.77	0.75 (1.0)	0.89 (0.87,0.91)
4-item	0.60	0.67 (1.1)	0.78 (0.73,0.82)	0.66	0.68 (1.1)	0.79 (0.74,0.83)	0.74	0.76 (1.2)	0.81 (0.77,0.85)	0.85	0.91 (1.3)	0.81 (0.77,0.84)

AD-only (n=505)

	Baseline to 1 Week Follow-up (n=440) [b]			Baseline to 3 Months Follow-up (n=416) [b]			Baseline to 6 Months Follow-up (n=434) [b]			Baseline to 12 Months Follow-up (n=440) [b]		
	Effect Size	Mean Change, Mean (SD)	ICC[c] (95% CI) of Mean Change	Effect Size	Mean Change, Mean (SD)	ICC[c] (95% CI) of Mean Change	Effect Size	Mean Change, Mean (SD)	ICC[c] (95% CI) of Mean Change	Effect Size	Mean Change, Mean (SD)	ICC[c] (95% CI) of Mean Change
82-item	0.28	0.22 (0.6)	------	0.42	0.34 (0.7)	------	0.49	0.41 (0.8)	------	0.64	0.52 (0.9)	------
55-item	0.27	0.21 (0.6)	0.97 (0.96,0.97)	0.41	0.32 (0.7)	0.97 (0.96,0.97)	0.47	0.38 (0.8)	0.97 (0.96,0.98)	0.59	0.48 (0.9)	0.98 (0.97,0.98)
34-item	0.24	0.20 (0.7)	0.94 (0.93,0.95)	0.35	0.29 (0.7)	0.95 (0.94,0.96)	0.40	0.33 (0.8)	0.95 (0.94,0.96)	0.51	0.43 (0.9)	0.96 (0.95,0.97)
15-item	0.20	0.18 (0.8)	0.90 (0.88,0.92)	0.33	0.29 (0.8)	0.91 (0.89,0.93)	0.35	0.32 (0.9)	0.90 (0.88,0.92)	0.44	0.41 (1.0)	0.93 (0.91,0.94)
9-item	0.23	0.21 (0.8)	0.84 (0.81,0.87)	0.33	0.30 (0.9)	0.87 (0.84,0.89)	0.36	0.35 (0.9)	0.87 (0.84,0.89)	0.45	0.44 (1.0)	0.90 (0.88,0.92)
4-item	0.26	0.27 (1.0)	0.76 (0.71,0.80)	0.35	0.37 (1.0)	0.79 (0.75,0.83)	0.38	0.41 (1.1)	0.81 (0.77,0.84)	0.47	0.51 (1.2)	0.84 (0.81,0.87)

[a] English-speakers and Spanish-speakers had similar results as shown in the Appendix 3 and 4.

[b]Missing values due to loss-to-follow up at various follow-up time points.

[c]All p-values for intraclass correlation coefficient (ICC) were <0.001.

Table 3.

Between-Group Effect Sizes and Differences of Mean Change Scores Over Time using Progressively Shorter ACP Engagement Survey Versions [a]

PREPARE versus AD-only	Baseline to 1 Week Follow-up		Baseline to 3 Months Follow-up		Baseline to 6 Months Follow-up		Baseline to 12 Months Follow-up	
	Effect Size of Differences	Differences of Mean Change, Mean (SD)[b]	Effect Size of Differences	Differences of Mean Change, Mean (SD)[b]	Effect Size of Differences	Differences of Mean Change, Mean (SD)[b]	Effect Size of Differences	Differences of Mean Change, Mean (SD)[b]
82-item	0.24	0.28 (0.7)	0.29	0.27 (0.8)	0.31	0.29 (0.8)	0.30	0.30 (0.9)
55-item	0.21	0.27 (0.6)	0.24	0.23 (0.7)	0.26	0.26 (0.8)	0.24	0.25 (0.9)
34-item	0.21	0.26 (0.7)	0.22	0.21 (0.8)	0.23	0.23 (0.8)	0.24	0.25 (0.9)
15-item	0.25	0.33 (0.8)	0.20	0.22 (0.8)	0.24	0.26 (0.9)	0.25	0.29 (1.0)
9-item	0.24	0.36 (0.9)	0.20	0.26 (0.9)	0.22	0.28 (0.9)	0.24	0.32 (1.0)
4-item	0.25	0.40 (1.0)	0.23	0.31 (1.1)	0.26	0.35 (1.1)	0.29	0.40 (1.2)

[a]English-speakers and Spanish-speakers had similar results as shown in the Appendix 5 and 6.

[b]T-tests for comparing differences of mean change between progressively shorter Survey versions and the original 82-item version all had non-significant p-values with Bonferroni adjustment for multiple comparisons at a significance level of 0.05, indicating no obvious differences among Survey versions.