



Published in final edited form as:

*Nicotine Tob Res.* 2006 December ; 8(Suppl 1): S49–S57.

## Internet- vs. telephone-administered questionnaires in a randomized trial of smoking cessation

**Amanda L. Graham, George D. Papandonatos, Beth C. Bock, Nathan K. Cobb, Arielle Baskin-Sommers, Raymond Niaura, and David B. Abrams**

Amanda L. Graham, Ph.D., Center for Alcohol and Addiction Studies, Brown Medical School, Providence, Rhode Island; Beth C. Bock, Ph.D., Centers for Behavioral and Preventive Medicine, Brown Medical School/The Miriam Hospital, Providence; George D. Papandonatos, Ph.D., Center for Statistical Sciences, Brown University, Providence; Nathan K. Cobb, M.D., Massachusetts General Hospital, Pulmonary and Critical Care Unit, Boston, Massachusetts; Arielle Baskin-Sommers, Brown University, Providence; Raymond Niaura, Ph.D., Brown Medical School/Butler Hospital, Providence; David B. Abrams, Ph.D., Office of Behavioral and Social Sciences Research, National Institutes of Health, Bethesda, Maryland.

### Abstract

The Internet offers a promising channel to conduct smoking cessation research. Among the advantages of Internet research are the ability to access large numbers of participants who might not otherwise participate in a cessation trial, and the ability to conduct research efficiently and cost-effectively. To leverage the opportunity of the Internet in clinical research, it is necessary to establish that measures of known validity used in research trials are reliable when administered via the Internet. To date, no published studies examine the psychometric properties of measures administered via the Internet to assess smoking variables and psychosocial constructs related to cessation (e.g., stress, social support, quit methods). The purpose of the present study was to examine the reliability of measures of previous quit methods, perceived stress, depression, social support for cessation, smoking temptations, alcohol use, perceived health status, and income when administered via the Internet. Participants in the present study were enrolled in a randomized controlled trial of the efficacy of Internet smoking cessation. Following baseline telephone assessment and randomization into the parent trial, participants were recruited to the reliability substudy. An email was sent 2 days after the telephone assessment with a link to the Internet survey and instructions to complete the survey that day. Of the 297 individuals invited to participate, 213 completed the survey within 1 week. Results indicate that the internal consistency and test–retest reliability of the measures examined are comparable when self-administered via the Internet or when interviewer-administered via telephone.

### Introduction

With approximately 70% of Americans having Internet access, the Internet offers a promising channel to conduct health behavior change research (Rainie & Horrigan, 2005). Among the advantages of conducting research on the Internet are the ability to reach large numbers of people who might not otherwise participate in clinical trials and the ability to conduct research efficiently and cost-effectively. However, to leverage the opportunities for clinical research

© 2006 Society for Research on Nicotine and Tobacco

Correspondence: Amanda L. Graham, Ph.D., Department of Oncology, Georgetown University Medical Center and Cancer Control Program, Lombardi Comprehensive Cancer Center, 3300 Whitehaven Street, NW, Suite 4100, Washington, DC 20007. Tel: 202 687 7905; Fax: 202 687 0305; Amanda\_Graham@brown.edu.

provided by the Internet, it is necessary to establish that measures known to be valid are reliable when administered via the Internet.

One concern about using the Internet for data collection in clinical trials is that data may be unreliable because researchers lack control over the environment in which data are collected. However, a growing body of evidence suggests that reliability and validity of data obtained using questionnaires administered via the Internet are generally consistent with results obtained through paper-and-pencil administered questionnaires (Davis, 1999; Gosling, Vazire, Srivastava, & John, 2004; Ritter, Lorig, Laurent, & Matthews, 2004). Cross-method consistencies have been demonstrated for numerous psychological and behavioral constructs including self-esteem (Robins, Trzeniewski, Tracy, Gosling, & Potter, 2002), self-trust (Buchanan & Smith, 1999; Foster, Campbell & Twenge, 2003), reaction time (McGraw, Tew, & Williams, 2000), depression (Lin et al., 2003), personality (Buchanan & Smith, 1999; Davis, 1999), and a variety of measures related to health status and health behaviors (Ritter et al., 2004).

Smoking cessation trials often include measures of co-occurring conditions and other factors related to cessation, including stress, depression, self-efficacy, social support, perceived health status, alcohol use, and previous quit methods. To date, there are no published studies comparing these measures when administered via the Internet and telephone. The present study examined the internal consistency and test–retest reliability of these measures in the context of a randomized controlled trial of the efficacy of Internet smoking cessation.

## Method

### Sample recruitment

Participants were enrolled in a parent study that is an ongoing randomized controlled trial of the efficacy of an Internet smoking cessation program ([www.quitnet.com](http://www.quitnet.com)) and telephone counseling. Recruitment leveraged QuitNet's top-ranked position on major search engines. Smokers in the United States were recruited into the parent trial if they used the terms *quit (ting) smoking* or *stop(ping) smoking* in a search engine query on one of four major search engines (AOL, MSN, Yahoo!, Google), and had not previously used QuitNet (no cookie detected). When an Internet user meeting these criteria clicked on a link to QuitNet in the results of a search engine query, they were shown an intercept page inviting them to participate in the parent study ("Do you want to quit smoking? Would you consider joining a Federally funded research study to evaluate QuitNet? All quit smoking services will be free as part of the study."). If they accepted, they were asked 10 eligibility screening questions (age, smoking rate, age of first puff, time to first cigarette after waking, number of quit attempts in past year, gender, race, education, zip code, and prior use of QuitNet website). Three questions determined preliminary eligibility (aged  $\geq 18$  years,  $\geq 5$  cigarettes/day, and no prior QuitNet use); remaining questions were used to characterize the largest possible denominator of potential study participants. Eligible participants were asked to provide online informed consent, their name, and contact information.

Within 48 hr, the participant was scheduled for the baseline telephone assessment, during which eligibility and informed consent were confirmed and a battery of assessments was administered via telephone interview. Participants were randomized to treatment, and then invited to participate in the Reliability Substudy. Subjects who consented to participate in the substudy were sent an e-mail 2 days after their telephone assessment that contained a link to the online survey. Each participant's unique study identification number was embedded into the link to the online survey so that responses could be joined with their telephone survey data. Participants were paid US\$15 for completing the online survey.

## Measures

The baseline telephone assessment for the parent trial consisted of measures assessing demographic, smoking, and psychosocial characteristics. Brief measures with known psychometric properties were selected to minimize respondent burden. The online reliability survey consisted of the following subset of the measures administered during the baseline telephone assessment.

**Quit methods**—Participants were asked whether they had ever used the following methods to quit smoking: Cold turkey, pamphlet or book, individual counseling, group counseling, nicotine patch, nicotine gum, nicotine nasal spray, nicotine lozenge, nicotine inhaler, bupropion, switching to chew tobacco or snuff, an Internet program (not QuitNet), telephone counseling, acupuncture, hypnosis, or any other method not listed.

**Smoking Temptations Questionnaire (Short Form)**—The 9-item version of the Smoking Temptations Questionnaire (Velicer, DiClemente, Rossi, & Prochaska, 1990) examined the degree to which various situations were tempting, a component of the construct of self-efficacy. Each item is rated on a 5-point scale ranging from 1=*not at all tempting* to 5=*extremely tempting*. The questionnaire can be scored to form a total score, as well as three subscales measuring temptations in positive affect or social situations, negative affect situations, and habitual or craving situations. Cronbach's alpha reliability coefficients for the three subscales are .84 (positive/social), .92 (negative affect), and .82 (habit/craving).

**Perceived Stress Scale—4-item**—The Perceived Stress Scale (PSS; Cohen, Kamarck, & Mermelstein, 1983) assesses the degree to which participants find their lives to be unpredictable and uncontrollable. Stress has been implicated in difficulty quitting smoking and relapse (Cohen et al., 1989). Each item is rated on a 5-point scale to indicate how frequently the individual has felt a particular way during the past month. Response options include *never* (0), *almost never* (1), *sometimes* (2), *fairly often* (3), and *very often* (4). A 4-item version has been validated, with Cronbach alpha reliability coefficients of .60 (Cohen & Williamson, 1988) to .72 (Cohen et al., 1983). Test-retest correlations range from .85 over 2 days in a college sample and .55 over 6 weeks in a smoking cessation sample (Cohen et al., 1983).

**Center for Epidemiological Studies-Depression Scale**—The 10-item Center for Epidemiological Studies—Depression (CES-D) Scale measures symptoms of current depression (Andresen, Malmgren, Carter, & Patrick, 1994). Scores on the CES-D have been positively associated with smoking prevalence and intensity, and failure to quit in representative samples of U.S. adults (Anda et al., 1990). Each item is rated on a 4-point scale to indicate the frequency of occurrence during the past week. Response options were modified to *less than 1 day* (0), *1–2 days* (1), *3–4 days* (2), and *5–7 days* (3). Test-retest correlations range from .21 to .84, with an overall correlation of .71, at an average time interval of 22 days (Andresen et al., 1994).

**Partner Interaction Questionnaire**—The Partner Interaction Questionnaire (PIQ; Cohen & Lichtenstein, 1990) consists of 20 items that measure specific behaviors of a spouse relevant to smoking cessation. Support involving cooperative behaviors (e.g., expressing pleasure at the smoker's efforts to quit) predicts successful quitting (Coppotelli & Orleans, 1985; Mermelstein, Lichtenstein, & McIntyre, 1983) and negative behaviors (e.g., nagging, criticizing) predict relapse (Cohen & Lichtenstein, 1990; Roski, Schmid, & Lando, 1996). The PIQ consists of 10 positive and 10 negative behaviors by the partner concerning the participant's smoking. Participants rate how frequently their partner will behave in a particular way from *never* (0), *almost never* (1), *sometimes* (2), *fairly often* (3), and *very often* (4). The PIQ has been modified to measure the receipt of specific behaviors from the person who follows

participants' quitting most closely, not just a partner (McMahon & Jason, 2000; Gruder et al., 1993). We selected the three positive and three negative items that loaded most strongly in a factor analysis (personal communication, R. J. Mermelstein, November 9, 2004). The three positive items were express pleasure at your efforts to quit, congratulate you for your decision to quit smoking, and express confidence in your ability to quit/remain quit. The three negative items were mention being bothered by smoke, ask you to quit smoking, and *criticize your smoking*. Cronbach alpha coefficients were .92 for the three-item positive subscale and .84 for the three-item negative subscale.

**Alcohol use**—Alcohol use is a common barrier to cessation (McClure, Wetter, de Moor, Cinciripini, & Gritz, 2002; Shiffman et al., 1996). Participants were asked if they drank alcohol at all. Those who said yes were asked to indicate how many days per week they drank alcohol on average, how many drinks they typically had, and the maximum number of drinks they had on one occasion during the past month. In addition, two questions asked, “In the last year, have you had more to drink than you meant to?” and, “In the last year, have you felt you wanted or needed to cut down on your drinking?” These items were adapted from a two-item screening measure to assess conjoint alcohol and drug use (Brown, Leonard, Saunders, & Papasouliotis, 2001). These items have high specificity (80%–90%) to detect current alcohol problems.

**Perceived health status, medical history and socio-demographics**—Participants were asked to rate their current health status on a 5-point scale from 1 (*excellent*) to 5 (*poor*). Participants were also asked if they had ever had smoking-related illness (*yes/no*). Total household income during the past year was assessed with eight response options (Table 3).

The survey questions were administered on nine separate screens. All questions were closed-ended, and needed to be completed in order to move to the next screen. Consistent with usability guidelines from the National Cancer Institute (usability.gov), a status bar (e.g., Page 3 of 9) appeared at the top of each page and the majority of questions appeared above the fold so that minimal scrolling was necessary. Multiple-choice questions appeared in a grid, with one question per row and one response per column; alternate questions were highlighted for ease of reading.

### Statistical analysis

The first set of analyses documents the recruitment process and describes the recruited sample. To examine the generalizability of the final sample, we compared survey completers with noncompleters on a range of demographic, smoking, and psychosocial variables. Frequency tables are used to summarize the categorical data, and parametric and nonparametric tests are employed to determine statistical significance.

The comparison of the Internet- and phone-administered measures is based on assessing the interrater reliability for these two survey methods and the presence of any systematic bias, as manifested by mean differences for continuous variables and prevalence differences for categorical variables. Table 1 provides sample means and standard deviations for all continuous variables. To allow for the presence of outliers in the data, *t*-tests for the mean difference between Internet- and phone-administered measures were supplemented with more robust nonparametric tests of location based on the Wilcoxon statistic. Additionally, standardized mean differences allowed us to use post-hoc power calculations to distinguish clinically from merely statistically significant results. This study has at least 80% power at the 5% level of significance to detect an effect size of  $\delta=.20$  using  $N=219$  subjects (the entire sample) and of  $\delta=.23$  using  $n=152$  subjects (those known to have been alcohol users). According to Cohen (1988), these are small effect sizes, which—while likely to be statistically significant

in our study—may have less practical importance than moderate effect sizes in the  $\Delta=.50$  range.

In Table 1, we also examined the test–retest reliability of all continuous variables across survey methods using the intraclass correlation coefficient (ICC) between a single rating obtained via the Internet and a single rating of the same measure obtained over the phone (formula 3.1 of Shrout & Fleiss, 1979). Interrater reliability above 80% is usually sought in method comparisons, with 70% considered an acceptable value.

Table 2 provides the prevalence of all dichotomous variables. Prevalence differences between the paired binary indicators contributed by each subject are tested using McNemar's test of marginal homogeneity. It is noteworthy that its power is driven entirely by the number of subjects with discordant reports ( $N_D$ ), rather than the total sample size. Therefore, in Table 2 the number of subjects ( $N$ ) contributing data on both surveys is supplemented by  $N_D$ , the actual sample size for each variable. Effect sizes for McNemar's test have been defined by Cohen (1988) as small for  $g=0.05$ , moderate for  $g=0.15$ , and large for  $g=0.25$ , where  $g$  is the absolute difference from 50% in the proportion of discordant pairs that endorse the Internet over the phone-administered measure. Detectability with 80% power at the 5% significance level requires that the observed number of discordant reports exceeds  $N_D=140, 79, \text{ and } 23$ , respectively.

Interrater agreement for binary variables is typically assessed using kappa coefficients (Cohen, 1960). Since their range is constrained by differences in prevalence between the dichotomous measures, caution should be exercised in their interpretation when the associated McNemar's test is significant (Cook, 1998). In the absence of prevalence differences, standard cutoffs for measuring agreement using kappa coefficients have been established by Landis and Koch (1977), which rate them as follows: .80–1.00=Almost Perfect, .60–.80=Substantial, .40–.60=Moderate, .20–.40=Fair, .00–.20=Slight, and <.00=Poor. Confidence intervals were based on the profile variance method of Lee and Tu (1994).

Prevalence differences between ordinal variables in Table 3 were assessed using Generalized Estimating Equation models for multinomial data, with a working exchangeable correlation matrix, as implemented in PROC GENMOD of SAS 8.2 (SAS Institute Inc., 2001). Extensions of kappa-type statistics to ordinal data proposed by Cohen (1968) require weights for the cells corresponding to partial agreement. In Table 3, we used linearly decreasing weights of the form  $1-(i-j)/(k-1)$ , where  $i$  and  $j$  refer to the row and column scores and  $k$  is the number of response categories. Health status was rated on a 5-point scale, whereas household income was scored using the category midpoints for all categories other than the last one, for which a sensitivity analysis was conducted by varying the midpoint from US\$125,000 to US\$150,000. Confidence intervals can be calculated using the normal approximation of Fleiss, Cohen, & Everitt (1969).

Finally, in Table 4 the internal consistency of several continuous scales is examined using Cronbach's alpha reliability coefficient (Cronbach, 1951) with 95% confidence intervals obtained according to van Zyl, Neudecker, and Nel (2000). Between-method comparisons were then performed by applying the methods outlined in Donner (1998) for comparing correlated alpha coefficients.

## Results

### Recruitment results and sample characteristics

The study was run for 16 weeks between June 2005 and September 2005. During that time, 297 individuals were invited to participate. Of those, 288 accepted (97%) and 221 completed

the survey for a final recruitment rate of 76.7%. Eight individuals completed the online survey more than 1 week after their telephone assessment and were excluded from analyses, resulting in a final sample of 213.

The majority of participants were female (65.3%), 79.8% were White, 11.7% Black, 4.2% Asian, 3.3% American Indian or Alaskan Native, and 2 individuals (<1%) were Native Hawaiian or other Pacific Islander. The average age of participants was 35.5 ( $SD=10.2$ ; range 18–70). Almost half of participants had completed 1–3 years of college (49.3%), followed by 31.9% with  $\geq 4$  years of college, 16% with a high school degree or GED, and 2.8% with less than a high school degree. Analysis of family income showed 27.1% earning less than US \$30,000 per year, 33.3% earning US\$30,000 to US\$50,000 per year, and 39.5% earning US \$50,000 or more.

The majority of participants (92.5%) reported that they were planning to quit in the next 30 days. On average, participants smoked 19.3 cigarettes per day ( $SD=9.9$ , range 5–60), had their first puff of a cigarette at age 14.2 ( $SD=3.4$ , range 7–29), and became daily smokers at age 17 ( $SD=3.6$ ; range 8–30). Participants had made an average of 2.9 quit attempts in the past year ( $SD=5.2$ ; range 0–50) and reported higher levels of desire to quit ( $M=8.95$ ,  $SD=1.54$ ) than confidence in quitting ( $M=6.24$ ,  $SD=2.29$ ). Average score on the Fagerström Test for Nicotine Dependence (Heatherton, Kozlowski, Frecker, & Fagerström, 1991) was 5.03 ( $SD=2.37$ ), with 44% of participants scoring 6 or above indicating a high level of nicotine dependence (Fagerström, Kunze, Schoberberger, et al., 1996). Analyses of body mass index indicated that 32.1% were overweight and 24.4% were obese according to standards of the Centers for Disease Control and Prevention (2004). Of those who reported current alcohol use (74.2%,  $n=152$ ), 60.8% indicated they had more to drink than they meant to in the past year and 21.5% indicated they wanted or needed to cut down on their drinking.

The majority of participants had used the Internet for more than 5 years (79.8%), accessed the Internet several times a day (76.1%), and used a high-speed Internet connection (85.8%). Almost half (44.1%) of participants used the Internet to communicate with other people through services like an Internet blog, online bulletin board, or instant messaging.

### Survey completers vs. noncompleters

Survey completers ( $n=213$ ) were compared with noncompleters ( $n=67$ ) to determine if there were group differences in demographic, smoking, and psychosocial characteristics. There were no differences on any of the variables tested, which included age, gender, race, education, employment, income, marital status, smoking rate, nicotine dependence score (Fagerström Test for Nicotine Dependence), desire or confidence in quitting, duration or frequency of Internet use.

### Means and descriptive data

Of the continuous variables examined in Table 1, the only variable showing differences of moderate effect size between the Internet- and phone administered version is the Negative Affect Situations subscale of Smoking Temptations, with the mean of the phone-administered measures 0.44 standard units higher than the mean of the Internet-administered version ( $p<.0001$ ). Positive Affect also tends to be higher when reported over the phone, although the observed difference of 0.17 standard units is considerably smaller, albeit statistically significant ( $p=.0193$ ). As a result, the Total Score of the Smoking Temptations scale shows an overall difference in the “small-to-moderate” range, with the two sample means 0.34 standard units apart ( $p<.0001$ ). Other measures such as the CES-D show statistically significant differences even though the observed effect sizes are small, a result of the ample power our sample size affords for detecting within-subject differences in continuous outcomes. This lack of

systematic bias between the two survey methods is accompanied by strong intraclass correlations, as depicted in Table 1.

Although the binary variables listed in Table 2 also show no significant differences between the two survey methods, the power of McNemar's test in the present study is quite low for all but large effect sizes because of the small number of discordant pairs ( $N_D < 23$  throughout); the only exception is self-reported smoking-related illness, whose prevalence is considerably higher when assessed over the phone (59.91% vs. 49.53%,  $N_D = 32$ ,  $p < .0001$ ). The ordinal variables listed in Table 3 (Income, Health Status) also show no significant differences in prevalence ( $p > .10$ ).

### Test–retest reliability and internal consistency results

The test–retest reliabilities for continuous variables measured in these two surveys are uniformly high (above 80%) for the Partner Interaction Questionnaire (PIQ) and the alcohol consumption measures, and moderately strong (in the 70%–80% range) for the Perceived Stress Scale (PSS) and CES-D. As seen in Table 1, the results are least satisfactory for the individual subscales of the Smoking Temptations Questionnaire scale, all of which fall below the 70% reliability threshold. Still, the overall scale (Total Score) is more reliable, as would be expected from a composite of three correlated subscales, with its ICC attaining exactly the 70% threshold.

Substantial test–retest agreement has also been obtained for the binary variables in Table 2, with kappa values above .70 for all but four variables assessing prior use of quit methods: Individual counseling, nicotine spray, Internet treatment, and telephone counseling. However, the precision of the reliability estimates is lower for binary than for continuous measures, and the 95% confidence intervals appear quite wide, allowing for the possibility that four additional variables show only moderate degrees of agreement between the two survey methods: Use of group counseling, nicotine inhaler, switching to chewing tobacco or snuff as methods to quit smoking, and report of ever having had a smoking-related illness.

In Table 3, we find evidence of substantial agreement for health status (weighted  $\kappa = .73$ ) and almost perfect agreement for the income measure (weighted  $\kappa = .93$ ).

Results for the income measure were not dependent on whether the midpoint of the highest income category used to construct the weights was changed from US\$125,000 to US\$150,000. Because of the informativeness of ordinal as opposed to binary measures, the confidence intervals are narrower which indicates improved precision in the estimates.

Finally, in Table 4 we evaluated the internal consistency of four scales and use them to compare the two survey methods. Cronbach's alpha reliability coefficients exceed 80% for CES-D under both methods, are in the 70% to 80% range for the PIQ and PSS scales, and only fall below 70% for the phone-administered version of the Smoking Temptations scale. Between-method comparisons show no statistically significant differences for PIQ, PSS, and CES-D. Between-survey differences for the Smoking Temptations scale are borderline significant at the 5% level; in this case, as in all others, the Internet version appears to have higher internal consistency than the phone version.

## Discussion

Results from this study indicate that the internal consistency and test–retest reliability of Internet-administered measures of stress, depression, smoking temptations, quit methods, alcohol use, health status, and income are comparable with telephone-administered measures. These results add to a growing body of evidence supporting the use of the Internet to gather

data in clinical trials. Given the often poor response rate to mailed surveys (Asch, Jedrzejewski & Christakis, 1997), the high cost of telephone surveys, and the limited generalizability of self-selected volunteers who participate in clinical settings, the Internet offers a reliable and cost efficient means of obtaining data from participants enrolled in a clinical trial.

There were no systematic differences between survey completers and noncompleters in this study. As expected, the sample was highly motivated to quit smoking and had a smoking history profile (e.g., age of onset, desire to quit) that was representative of the smoking population who are proactively seeking treatment. Although it is common for females to enroll in a clinical trial at a higher rate than males, the Internet-based sample may have even more females and may be slightly younger than participants in typical clinical studies. There was a small but reasonable representation of minorities based on general Internet demographics, but as expected, the sample was primarily White and familiar with the Internet. Although results may be limited to this group, it is also reasonable to expect that this may be the group that would participate in Internet trials anyway. The moderate to high degree of nicotine dependence and presence of co-occurring conditions (stress, depression, alcohol use, overweight) seen in this sample may make cessation more difficult. Higher scores on the CES-D seen in the Internet version may indicate that participants are more comfortable reporting negative affect via the Internet than in a phone interview.

Reports of ever having had a smoking-related illness were more common in the telephone assessment than via the Internet. This is likely related to the wording of the question and the confusion of some participants regarding what constitutes a smoking-related illness. During the telephone assessment, research staff were able to clarify that this question included any health condition the participant believed was caused by their smoking, perhaps explaining the higher endorsement than in the Internet version. Income often is thought to be a sensitive question that participants might not feel comfortable answering; however, our findings indicate this is not the case and that participants provided the same information online as they did via telephone.

Test-retest reliability coefficients were lowest for the Smoking Temptations Scale, with ICC coefficients ranging from .66 to .68. This finding is likely related to the dynamic nature of cigarette craving, especially in a sample of smokers who had recently enrolled in a cessation trial (Velicer et al., 1990). Differences in internal consistency for this measure may also be related to difficulty understanding the question and/or the rating scale when administered on the telephone. We have found that this measure required repeated explanation during telephone administration. It may be that having the question and response options visually available during Internet administration facilitates response fidelity.

The study is limited by the necessity to administer brief forms of the measures to reduce response burden as well as by the self-selection biases inherent in any recruitment method. However, for clinical trials that will be conducted over the Internet, the results are encouraging. The sample appears to be quite representative of smokers who are interested in quitting. Future studies will need to replicate and extend the results reported here to other measures (including measures of cessation and verification of cessation in treatment outcomes) and to other recruitment methods (e.g., proactive vs. reactive recruitment; recruited via mail, telephone, print, radio or mass media). For cessation trials that examine individual differences and are interested in mechanisms of action (e.g., mediation/moderation models; the role of co-morbidity in treatment), the present study provides some of the first evidence that we know of to date that these measures are psychometrically sound and can be used in Internet-based smoking cessation research trials.

## Acknowledgments

This research was supported by the National Cancer Institute (5R01CA104836-02). We acknowledge and thank Lan Jiang, M.S., Crystal Davis, MPH, and Pearl Zakrofsky, B.A. for their assistance on this project.

## References

- Anda RF, Williamson DF, Escobedo LG, Mast EE, Giovino G, Remington PL. Depression and the dynamics of smoking: A national perspective. *Journal of the American Medical Association* 1990;264:1541–1545. [PubMed: 2395193]
- Andresen EM, Malmgren JA, Carter WB, Patrick DL. Screening for depression in well older adults: Evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression Scale). *American Journal of Preventive Medicine* 1994;10(2):77–84. [PubMed: 8037935]
- Asch DA, Jedrzejewski KM, Christakis NA. Response rates to mail surveys published in medical journals. *Journal of Clinical Epidemiology* 1997;50:1129–1136. [PubMed: 9368521]
- Buchanan T, Smith JL. Using the Internet for psychological research: Personality testing on the World-Wide-Web. *British Journal of Psychology* 1999;90:125–144. [PubMed: 10085550]
- Brown RL, Leonard T, Saunders LA, Papasouliotis O. A two-item conjoint screen for alcohol and other drug problems. *Journal of the American Board of Family Practitioners* 2001;14:95–106.
- Centers for Disease Control and Prevention. Body Mass Index: BMI Calculator. [Retrieved October 15, 2005]. 2004 from: <http://www.cdc.gov/nccdphp/dnpa/bmi/calc-bmi.htm>
- Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 1960;20:37–46.
- Cohen J. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin* 1968;70:213–20. [PubMed: 19673146]
- Cohen, J. *Statistical power analysis for the behavioral sciences*. 2nd Ed. L. Erlbaum; Hillsdale, NJ: 1988.
- Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *Journal of Health and Social Behavior* 1983;24(4):385–96. [PubMed: 6668417]
- Cohen S, Lichtenstein E. Partner behaviors that support quitting smoking. *Journal of Consulting and Clinical Psychology* 1990;58:304–309. [PubMed: 2365893]
- Cohen S, Lichtenstein E, Prochaska JO, Rossi JS, Gritz ER, Carr CR, Orleans CT, Schoenbach VJ, Biener L, Abrams DB, DiClemente C, Curry S, Marlatt GA, Cummings KM, Emont SL, Giovino G, Ossip-Klein D. Debunking myths about self-quitting: Evidence from 10 prospective studies of persons who attempt to quit smoking by themselves. *American Psychologist* 1989;44:1355–1365. [PubMed: 2589730]
- Cohen, S.; Williamson, G. Perceived stress in a probability sample of the United States. In: Spacapan, S.; Oskamp, S., editors. *The social psychology of health: Claremont Symposium on applied social psychology*. Sage; Newbury Park, CA: 1988.
- Cook, RJ. Kappa and its dependence on marginal rates. In: Armitage, TP.; Colton, editors. *The encyclopedia of biostatistics*. Wiley; New York: 1998. p. 2160-2166.
- Coppotelli H, Orleans C. Partner support and other determinants of smoking cessation maintenance among women. *Journal of Consulting and Clinical Psychology* 1985;53:455–460. [PubMed: 4031200]
- Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16:297–333.
- Davis RN. Web-based administration of a personality questionnaire: Comparison with traditional methods. *Behavior Research Methods, Instruments & Computers* 1999;31(4):572–577.
- Donner A. Sample size requirements for the comparison of two or more coefficients of inter-observer agreement. *Statistics in Medicine* 1998;17(10):1157–1168. [PubMed: 9618775]
- Fagerström KO, Kunze M, Schoberberger R, Breslau N, Hughes JR, Hurt RD, Puska P, Ramstrom L, Zatonski W. Nicotine dependence versus smoking prevalence: comparison among countries and categories of smokers. *Tobacco Control* 1996;5(1):52–56. [PubMed: 8795860]
- Fleiss JL, Cohen J, Everitt BS. Large sample standard errors of kappa and weighted kappa. *Psychological Bulletin* 1969;72:323–327.

- Foster J, Campbell W, Twenge J. Individual differences in narcissism: Inflated self-views across the lifespan and around the World. *Journal of Research in Personality* 2003;37:469–486.
- Gosling SD, Vazire S, Srivastava S, John OP. Should we trust Web-based studies? A comparative analysis of six pre-conceptions about Internet questionnaires. *American Psychologist* 2004;59:93–104. [PubMed: 14992636]
- Gruder CL, Mermelstein RJ, Kirkendol S, Hedeker D, Wong SC, Schreckengost J, Warnecke RB, Burzette R, Miller TQ. Effects of social support and relapse prevention training as adjuncts to a televised smoking-cessation intervention. *Journal of Consulting and Clinical Psychology* 1993;61:113–120. [PubMed: 8450096]
- Heatherton TF, Kozlowski LT, Frecker RC, Fagerström KO. The Fagerström Test for Nicotine Dependence: A revision of the Fagerström Tolerance Questionnaire. *British Journal of Addictions* 1991;86:1119–27.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74. [PubMed: 843571]
- Lee JJ, Tu ZN. A better confidence for kappa on measuring agreement between two raters with binary outcomes. *Journal of Computational and Graphical Statistics* 1994;3:301–321.
- Lin CC, Li YC, Bai YM, Tsai SJ, Hsiao MC, Wu CH, Liu CY, Chen JY. The validity of an Internet-based self-assessment program for depression. *American Medical Informatics Association Annual Symposium Proceedings* 2003:911.
- McClure JB, Wetter DW, de Moor C, Cinciripini PM, Gritz ER. The relation between alcohol consumption and smoking abstinence: Results from the Working Well Trial. *Addictive Behaviors* 2002;27:367–379. [PubMed: 12125663]
- McGraw K, Tew M, Williams J. The integrity of Web-delivered experiments: Can you trust the data? *Psychological Science* 2000;1:502–506. [PubMed: 11202497]
- McMahon SD, Jason LA. Social support in a worksite smoking intervention. A test of theoretical models. *Behavior Modification* 2000;24:184–201. [PubMed: 10804679]
- Mermelstein R, Lichtenstein E, McIntyre K. Partner support and relapse in smoking-cessation programs. *Journal of Consulting and Clinical Psychology* 1983;51:465–466. [PubMed: 6863711]
- Rainie, L.; Horrigan, J. Internet evolution. *Pew Internet and American Life Project*; 2005.
- Ritter P, Lorig K, Laurent D, Matthews K. Internet versus mailed questionnaires: A randomized comparison. *Journal of Medical Internet Research* 2004;6:e29. [PubMed: 15471755]
- Robins R, Trzeniewski K, Tracy J, Gosling S, Potter J. Global self-esteem across the life span. *Psychology and Aging* 2002;17:423–434. [PubMed: 12243384]
- Roski J, Schmid LA, Lando HA. Long-term associations of helpful and harmful spousal behaviors with smoking cessation. *Addictive Behaviors* 1996;21:173–185. [PubMed: 8730519]
- SAS Institute Inc.. *SAS/STAT Software: Changes and Enhancements*. Cary, NC: 2001. Release 8.2
- Shiffman S, Hickcox M, Paty JA, Gnys M, Kassel JD, Richards TJ. Progression from a smoking lapse to relapse: Prediction from abstinence violation effects, nicotine dependence, and lapse characteristics. *Journal of Consulting and Clinical Psychology* 1996;64:993–1002. [PubMed: 8916628]
- Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin* 1979;86:420–428. [PubMed: 18839484]
- van Zyl JM, Neudecker H, Nel DG. On the distribution of the maximum likelihood estimator of Cronbach's alpha. *Psychometrika* 2000;65:271–280.
- Velicer WF, DiClemente CC, Rossi JS, Prochaska JO. Relapse situations and self-efficacy: An integrative model. *Addictive Behaviors* 1990;15:271–283. [PubMed: 2378287]

Table 1

Tests of mean differences for continuous variables.

Variable name	N/n	Phone administered		Internet administered		Phone/Internet difference		Standardized mean difference	t-test p-value	Wilcoxon test p-value	Interrater reliability ICC 95% CI
		Mean	SD	Mean	SD	Mean	SD				
Smoking Temptations											
Total score	213	3.90	0.45	3.77	0.52	0.13	0.38	0.34	<.0001	<.0001	.70 (.62-.76)
Positive affect or social situations	213	3.67	0.68	3.58	0.79	0.10	0.60	0.17	.0193	.0187	.67 (.59-.74)
Negative affect situations	213	4.51	0.58	4.28	0.66	0.23	0.52	0.44	<.0001	<.0001	.66 (.57-.73)
Habitual or craving situations	213	3.51	0.72	3.45	0.77	0.07	0.59	0.12	.0975	.1205	.68 (.60-.75)
Partner Interaction Q											
Total score	208	15.62	5.26	15.14	5.51	0.48	2.60	0.18	.0088	.0032	.88 (.85-.91)
Positive	209	9.65	2.42	9.49	2.64	0.16	1.52	0.11	.1233	.0840	.82 (.77-.86)
Negative	208	5.96	4.29	5.64	4.45	0.32	2.03	0.16	.0254	.0132	.89 (.86-.92)
Cohen Perceived Stress Scale	213	5.92	2.88	6.21	3.06	-0.29	2.12	-0.14	.0500	.0179	.74 (.68-.80)
CES-D	213	8.79	5.53	9.38	5.96	-0.59	3.68	-0.16	.0210	.0047	.79 (.74-.84)
No. drinking days per week	151	1.87	1.72	1.87	1.62	0.00	0.51	-0.00	.9426	.9220	.95 (.94-.97)
No. drinks on typical day	151	3.05	2.09	3.06	2.13	-0.00	1.19	-0.00	.9725	.4210	.84 (.79-.88)
Max no. drinks on single occasion	151	5.83	5.43	5.87	4.43	-0.03	1.94	-0.02	.8342	.0499	.92 (.90-.94)

ICC, Intraclass correlation; CES-D, Center for Epidemiological Studies—Depression Scale.

**Table 2**

Tests of prevalence differences and degree of agreement for binary variables.

Variable name	N/n	N <sub>D</sub>	Phone administered		Internet administered		Phone-Internet difference		McNemar's Test		Degree of agreement	
			Yes (%)	Yes (%)	Yes (%)	Yes (%)	Y	Y	p-value	κ	κ	95% CI
Quit methods (ever used)												
Cold turkey	213	22	80.28	83.10	-2.82	.2863	.63	(.48-.75)				
Pamphlet or booklet	213	21	21.60	23.94	-2.35	.3833	.70	(.56-.79)				
Individual counseling	213	9	4.69	4.23	0.47	1.0000	.42	(.17-.67)				
Group counseling	213	9	7.51	7.98	-0.47	1.0000	.73	(.51-.86)				
Nicotine patch	213	8	53.05	52.11	0.94	.7266	.93	(.86-.96)				
Nicotine gum	213	8	37.09	35.21	1.88	.2891	.90	(.82-.95)				
Nicotine spray	213	2	0.47	1.41	-0.94	.5000	.50	(.09-.86)				
Nicotine lozenge	213	5	10.80	9.39	1.41	.3750	.88	(.72-.95)				
Nicotine inhaler	213	6	6.57	6.57	0.00	1.0000	.77	(.54-.89)				
Bupropion	213	7	35.21	36.62	-1.41	.4531	.92	(.85-.96)				
Switch to chew tobacco/snuff	213	6	8.45	8.45	0.00	1.0000	.81	(.61-.91)				
Internet	213	16	6.57	6.57	0.00	1.0000	.39	(.17-.61)				
Telephone counseling	213	7	3.76	4.23	-0.47	1.0000	.57	(.28-.79)				
Acupuncture	213	1	6.57	7.04	-0.47	1.0000	.96	(.80-.99)				
Hypnosis	213	2	15.02	15.96	-0.94	.5000	.95	(.85-.98)				
Ever had smoking related illness?	212	32	59.91	49.53	10.38	.0001	.71	(.60-.79)				
Do you drink alcohol?	213	8	74.18	74.18	0.00	1.0000	.88	(.79-.94)				
Drink more than meant to	154	8	61.69	60.39	1.30	.7266	.86	(.76-.93)				
Wanted/needed to cut down	154	8	22.08	20.78	1.30	.7266	.84	(.70-.92)				

N<sub>D</sub>, Number of discordant pairs.

Table 3

Test of prevalence differences and degree of agreement for ordinal variables.

Variable name	N/n	Phone administered		Internet administered		Location test		Degree of agreement	
		Yes (%)	Yes (%)	Yes (%)	Yes (%)	p-value	$\kappa^b$	95% CI	
Income <sup>a</sup>	210					.1382	.93	(.85–1.00)	
Less than US\$10,000		1.90	2.38						
US\$10,000 to US\$20,000		9.05	9.05						
US\$20,000 to US\$30,000		16.19	17.62						
US\$30,000 to US\$40,000		17.62	13.81						
US\$40,000 to US\$50,000		15.71	18.10						
US\$50,000 to US\$75,000		20.48	20.00						
US\$75,000 to US\$100,000		9.52	10.00						
US\$100,000 or more		9.52	9.05						
Health status	212					.1109	.73	(.64–.82)	
Excellent		7.55	6.60						
Very good		39.62	40.57						
Good		35.38	38.21						
Fair		15.57	13.68						
Poor		1.89	0.94						

<sup>a</sup>Income scored at category midpoint; US\$125,000 used for last category.

<sup>b</sup>Weighted kappa using absolute difference between category scores to define a distance measure.

**Table 4**

Internal consistency comparisons for Internet- versus interviewer-administered measures.

Variable name	Phone administered (N=213)		Internet administered (N=213)		Test of difference
	C( $\alpha$ ) <sup>a</sup>	95% CI	C( $\alpha$ ) <sup>a</sup>	95% CI	
Smoking Temptations	.63	(.56-.70)	.70	(.64-.76)	.05
Partner Interaction Questionnaire	.75	(.70-.80)	.78	(.74-.83)	.14
Cohen Perceived Stress Scale	.72	(.65-.78)	.78	(.72-.82)	.21
CES-D	.80	(.76-.84)	.85	(.82-.88)	.07

Note: *p*-values represent two-tailed test of the difference of the two correlated reliability coefficients.

<sup>a</sup> C( $\alpha$ )=Cronbach's alpha based on items standardized to zero mean and unit variance. CES-D, Center for Epidemiological Studies—Depression Scale.