Why Do People Report Better Health by Phone Than by Mail?

Noel T. Brewer, PhD, William K. Hallman, PhD,† Nancy Fiedler, PhD,‡ and Howard M. Kipen, MD‡*

Context: Past research shows that fewer health symptoms are reported by phone than by mail.

Objectives: We sought to examine whether interview mode-dependent differences in health symptom reporting are the result of socially desirable responding or to expending less cognitive effort when formulating responses, a behavior known as satisficing.

Design: Participants were randomly assigned to telephone interview only or to mail interview followed 2 weeks later by telephone interview.

Setting & Participants: Participants were American veterans from the Gulf War Registry (n = 719).

Main Outcome Measures: Our main outcome measure was the number of mild, moderate, or severe symptoms reported (of 48 possible).

Results: Veterans reported an average of 5 more symptoms via mail than via telephone, F (1, 709) = 32.50, $P < 0.001$. The difference was mainly the result of symptoms reported by mail as mild but not reported at all by phone. Veterans with higher social desirability scores reported fewer symptoms by phone and mail, F (1, 709) = 10.11, $P = 0.001$, but social desirability scores did not interact with interview mode. Furthermore, embarrassing symptoms such as genital complaints were no less likely to be reported by phone.

Conclusions: Reporting of better health in phone surveys is the result of fewer mild symptoms reports but not of socially desirable responding. The findings are consistent with phone interviews encouraging satisficing by limiting the recall of less severe health states. Researchers should handle mild symptom reports with some skepticism.

Key Words: social desirability, symptom reporting, interview mode, satisficing

(*Med Care* 2004;42: 875–883)

Symptoms play an important role in understanding patients and their health behaviors. Clinicians commonly ask patients for reports of pain and other symptoms of illness. Medical and psychological researchers rely on symptom reports both as interesting in their own right and as predictors of other important outcomes, such as medication compliance.[1,2] Moreover, self-reported symptoms are the only evidence that clinicians and researchers have of certain illnesses such as chronic fatigue syndrome, Gulf War illness,[3,4] and chronic Lyme disease.[5]

Despite the best of intentions, patients' symptom reporting can be incomplete or biased. Patients may misunderstand the question asked, fail to retrieve what they know, not fully integrate relevant information into a coherent judgment, or improperly use the response scale.[6,7] One serious problem is that interview mode affects the number of health symptoms reported. Mailed surveys routinely yield more symptom reports than phone (or in-person) interviews.[8,9] Because phone and in-person interviews show no reliable differences in symptom reports,[10] the present research focused on mail-phone differences.

Explanations for interview mode differences emphasize the role of socially desirable responding, a tendency to withhold stigmatized information about oneself. The argument goes that people withhold symptom reports on the phone because self-consciousness causes easily embarrassed people to offer replies that are more flattering or socially desirable.[10] Socially desirable responding can be conceptualized as topic-based or personality-based. The former suggests that some topics inhibit candid answers. Classic examples are that questions on having voted or having a library card elicit overreporting,[11] but questions on bankruptcy or drunk driving

From the *Department of Health Behavior and Health Education, School of Public Health, University of North Carolina, Chapel Hill, North Carolina; †Food Policy Institute, Rutgers University, New Brunswick, New Jersey; and the ‡Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, Piscataway, New Jersey.

Supported by grants awarded to Dr. Kipen by the Department of Veterans Affairs and the Centers for Disease Control and Prevention (Grant #U50/CCU214463-01). Drs. Kipen and Fiedler also were supported by NIEHS Center Grant ES05022.

Portions of this paper were presented at the 2003 Annual Conference of the Society for Behavioral Medicine, Salt Lake City, UT.

Reprints: Noel T. Brewer, PhD, Department of Health Behavior and Health Education, UNC School of Public Health, CB#7440 Rosenau Hall, Chapel Hill, NC 27599. E-mail: ntb@unc.edu.

ISSN: 0025-7079/04/4209-0875

elicit underreporting.[12] This sort of finding is then used to infer that the same goes for health reports. Although such a conclusion is tentatively supported by research showing that socially sensitive questions and less socially desirable topics elicit less candid responses on the phone,[9,12–14] the research does not identify the person-level variables responsible for the effect.

The latter research links socially desirable responding to a personality trait characterized by lower self-reports of unflattering behaviors periodically engaged in by most people.[15,16] Scales measuring this trait tally denials to questions like these: "I like to gossip at times," and "I sometimes feel resentful when I don't get my way."[17] We will refer to this person-level tendency as the social desirability score. Other terms for trait social desirability (ie, the desire to portray oneself in a flattering light) are defensiveness and need for approval. For the purposes of this article, the term social desirability will refer to the personality trait and socially desirable responding to its influences. Those with higher social desirability scores generally report fewer physical and psychological symptoms (Brewer NT, Hallman WK, Fiedler N, et al. submitted for publication).[18–24] For example, Koller et al[22] found that higher social desirability scores predicted fewer symptom reports among postsurgery cancer patients. Socially desirable responding appears to reflect more than just human fallibility[23]and may even have positive implications for psychological adjustment.[24,25]

The correspondence between social desirability scores and symptom reports suggests a direct way to test the social desirability explanation for interview mode differences. Past research has shown similar social desirability scores in paper- and computer-administered surveys but higher scores in face-to-face interviews.[26] However, the studies examined *mean* levels of socially desirable responding. They did not examine whether the *relationship* of social desirability scores to symptom reporting differed by interview mode. In other words, social desirability scores are related to interview mode and to symptom reporting, but there may be no interaction. To our knowledge, no one has yet empirically tested the interaction (but see[27]).

An alternative hypothesis concerns people's tendency to reduce the cognitive burden placed on them during phone interviews. Because formulating a full and complete answer may be too effortful, they may offer a compromise response that they think is "good enough," a behavior known as satisficing.[28,29] People offer responses that reflect only a portion of what they know (ie, weak satisficing) when a full search of their memory for all they know is too burdensome. Some people may go even further and report none of what they know (ie, strong satisficing) by using extreme tactics such as answering the first of a series of questions truthfully and then repeating that answer for the remainder. The demands inherent to phone interviews may encourage satisficing more than mail surveys do. In mail surveys, people proceed at their own pace and can see both the question and all response options. They have more time to search their memory for relevant information, formulate a response, and map it onto the responses provided. Phone interviews provide a contrasting experience for interviewees who must hold in memory the exact question and response options and who no longer control the pace of the survey. Waiting while questions are asked and answering branching follow-up questions may make phone interviews seem more tedious and time consuming.

The present study looked at 2 competing explanations for interview mode differences in symptom reporting. The first concerns response editing. Responders with higher social desirability scores (already known to report relatively fewer symptoms) may report even fewer symptoms by phone than by mail (presumably because the presence of the interviewer is more salient). In statistical terms, the social desirability hypothesis would be supported by an interaction of social desirability score and interview mode in predicting symptoms. The second explanation concerns the effort put into reporting symptoms. Severe health symptoms should be more accessible in memory and, as a result, easier to recall.[30] The result of the less effortful recall should be less satisficing for reports of severe symptoms. Satisficing could also happen if symptoms were recalled but selectively not reported to save time. A satisficing account would be supported by an interaction of interview mode and symptom severity. In particular, the account suggests that symptoms reported by mail would be less easily recalled on the phone and thus be less likely to be reported.

METHOD

Participants

Participants were US military veterans from the Gulf War Health Registry. The veterans were randomly selected from the Registry and interviewed previously in 1995.[31] Veterans who completed that survey formed the sample for the present study that was conducted in 2000. Seventy of the veterans who completed the 1995 interview were later identified as ineligible for further study due to incarceration, physical incapacitation, not being eligible for the Gulf War Registry, or death.

Procedure

Participants were interviewed by telephone with the assistance of computer-interviewing software. Just more than half of them were randomly assigned to and completed an additional mail survey 2 weeks in advance of their phone interview. Thus, participants were either interviewed once by phone only or they were interviewed once by mail and again by telephone. The phone survey contained all measures

reported here, including health symptoms and social desirability, and the mail survey contained only the symptom questions. To make the 2 surveys as similar as possible, the symptom questions started off both surveys, just after an initial summary health question. The authors' institutional review boards approved the study protocol.

Measures

The mail and telephone survey assessed the 48 health symptoms shown in Table 1. Telephone participants were instructed, "For each symptom, please tell me if in the last *six months*, you have had *persistent* or *recurring* problems with it. If yes, please tell me whether your problem was mild, moderate, or severe." In practice, phone participants usually made a single response and did not require prompting. The mail survey instructions read, "In the last 6 months, have you had persistent or recurring problems with..." and a 4-point response scale was provided for each symptom, labeled "none," "mild," "moderate," and "severe" with a vertical bar separating the "none" and severity options. Thus, in both surveys, participants were encouraged to first consider the symptom and its severity and then to respond using equivalent scales.

An index was created by counting the total number of symptoms each person endorsed as mild. Indices of moderate and severe symptom reports were created in a similar manner. For each symptom, a severity index was calculated by coding "no" as "0," "mild" as "1," "moderate" as "2," and "severe" as "3." A difference score was then calculated for each symptom. Some mail surveys were returned with missing or multiple responses to questions. Missing responses were coded as "none," and multiple responses were coded as the less severe of the 2.

Social desirability was assessed using 10 items drawn from the 33-item Marlowe-Crowne social desirability scale.[17] The shortened form has been shown to have adequate internal consistency ($\alpha = 0.63$) and to be well correlated to the full scale ($r = 0.85$).[32] Participants received 1 point for each socially desirable answer for a maximum possible score of 10 points. The scale's internal consistency in the present study was $\alpha = 0.62$.

Demographic variables assessed included age, gender, ethnicity, education level, and military rank. Ethnicity was dichotomized to create 2 groups, white and "other" ethnicity, because of the relatively small size of the remaining ethnic groups. To gauge better whether random assignment was successful, we compared physical function (SF-36)[33] and psychological distress across interview conditions (Brief Symptom Inventory).[34]

Data Analysis

The experimental design allows the 2 hypotheses about interview mode effects to be examined both between-subjects and within-subjects. The comparisons are described below but an inspection of Table 2 may help the reader to visualize them. The first analysis examined differences in symptom reports between the groups interviewed by mail or only by phone (ie, between bracketed means in the first row and last column of Table 1). A 2 (interview mode) $\times$ 3 (symptom severity) analysis of covariance (ANCOVA) for mixed designs was conducted predicting symptoms. The first variable was interview mode (ie, mail or phone). It was entered as a between-subjects categorical variable. The second variable was symptom severity (ie, mild, moderate, or severe). It was entered as a within-subjects categorical variable. The third variable, social desirability score, was mean-centered and entered as a continuous predictor. All 2 and 3-way interactions involving these 3 variables were included in the model. The model controlled for age, gender, military rank, ethnicity, and education as well as testing the interaction of education and interview mode. (Past research on satisficing has examined whether education, as a proxy for cognitive sophistication, attenuates satisficing and found mixed results. We include the relevant analyses here for the sake of completeness.) The second analysis examined symptom reports within the mail-phone interview group (ie, the bracketed means in the bottom row and last column of Table 1). The ANCOVA was repeated with the change that interview mode was within-subjects. We also examined the change in symptom reporting separately for each symptom using *t* tests, as well as for 5 clusters of symptoms that were suggested by a previously reported factor analysis.[31] A final analysis addressed an issue related to, but not central to, our theoretical concerns. It examined whether completing a recent mail survey biased later phone responses by comparing the phone interview responses of the 2 groups (ie, comparing bracketed means in the top and bottom rows of Table 1). The analysis was essentially a replication of the first ANCOVA in which the interview mode variable was replaced by a between-subjects variable representing recent-prior-interview (present or absent).

RESULTS

The experiment's design provides 2 ways to look at interview mode effects. We can look at differences in symptom reports between 1 group's mail reports and the other's phone reports. For some of these participants, we can pair their mail and telephone symptom reports to examine within-subjects differences.

Response Rates

Figure 1 shows the completion rates for the present study. Participants who were successfully interviewed (n = 719) were primarily white (81%), male (92%), low socioeconomic status (88% enlisted; mean 12.3 years edu-

TABLE 1. Count of Participants with Changes in the Severity of Symptom Reports Between Telephone and Mail

	Degree of Change in Reported Symptom Severity								
	−3	−2	−1	0	+1	+2	+3	Average	t
Mood-memory-fatigue symptoms	1	8	50	242	75	18	4	0.14	6.61†
41† Prolonged fatigue or feeling of illness after mild exercise	1	5	29	241	77	30	15	0.35	7.58†
42 Unexplained weakness	1	2	31	241	89	29	5	0.31	7.69†
39 Sleeping more than usual	6	8	27	261	58	31	7	0.20	4.32†
40 Fatigue (not due to exercise)	1	8	56	215	92	22	4	0.18	4.21†
44 Sudden mood changes	0	8	50	247	73	16	4	0.13	3.23†
48 Feeling depressed or blue	0	7	47	246	85	12	1	0.13	3.51†
37 Feeling sickly	0	5	52	257	65	15	4	0.11	2.98‡
35 Difficulty concentrating	1	5	66	229	85	11	1	0.08	2.04‡
43 Feeling anxious or upset	2	12	53	240	68	19	4	0.09	2.03‡
34 Losing your balance or feeling dizzy	0	8	53	256	72	9	0	0.05	1.51
36 Difficulty remembering things	1	7	56	251	73	10	0	0.05	1.40
38 Unrefreshing sleep (waking up tired)	1	15	84	219	64	14	1	−0.06	−1.31
Muscluloskeletal symptoms	2	18	56	226	72	22	4	0.08	3.03*§
25 Muscle aches or cramps	2	15	48	219	84	25	5	0.16	3.50§
28 Pain in your arms or legs	0	12	49	233	76	23	5	0.16	3.69§
16 Back problems	1	11	52	230	74	27	3	0.15	3.43§
29 Pain in more than one joint without swelling or redness	2	16	62	227	63	22	6	0.06	1.34
27 Numbness or tingling sensations	3	21	60	220	77	15	2	0.01	0.11
15 Arm, hands, or shoulder	1	30	64	225	58	19	1	−0.07	−1.53
Gastrointestinal symptoms	1	8	38	272	60	17	2	0.11	5.52§
19 Stomach or digestive system	1	12	53	227	76	24	5	0.15	3.30§
23 Abdominal pain	0	13	45	247	62	31	0	0.13	3.19‡
20 Diarrhea	0	3	47	266	66	14	2	0.12	3.38§
21 Constipation	0	4	24	303	55	11	1	0.12	4.02§
17 Nausea	1	9	24	293	53	16	2	0.12	3.26§
22 Abdominal gas	3	11	55	234	74	17	4	0.09	1.96
18 Vomiting	0	1	19	337	33	6	2	0.08	3.06‡
Throat-breathing symptoms	1	6	29	297	49	14	2	0.09	4.56§
7 Ability to taste	1	4	18	317	44	13	1	0.11	3.67§
8 Difficulty swallowing	0	3	24	313	46	9	3	0.11	3.63§
9 Throat problems	1	7	32	289	47	20	2	0.11	3.04§
11 Difficulty breathing	3	9	43	270	58	14	1	0.05	1.27
Other symptoms	1	10	32	272	57	22	3	0.14	9.03§
2 Eyes or vision	0	3	36	244	81	32	2	0.27	6.95§
12 Coughing	0	9	27	243	89	27	3	0.27	6.65‡
1 Headache	0	7	43	236	74	34	4	0.24	5.67§
4 Nose or sinuses	0	12	43	232	71	33	7	0.23	4.96§
6 Mouth, teeth, or gums	2	7	30	254	66	37	2	0.24	5.68§
3 Ears or hearing	1	9	36	239	78	35	0	0.23	5.44§
13 Chest discomfort or pain	1	10	42	243	81	16	5	0.16	3.82‡
26 Sexual or genital problems	1	10	24	287	49	18	9	0.16	3.99§
31 Hair	4	5	15	310	37	18	9	0.16	3.97§
14 Irregular heartbeat	0	4	19	310	48	16	1	0.14	4.60‡
5 Extra sensitivity to everyday chemicals like cleaners, paints, perfumes, or soaps	1	12	39	268	46	28	4	0.12	2.84‡

(*Continued*)

TABLE 1.(*Continued*)

	Degree of Change in Reported Symptom Severity								
	−3	−2	−1	0	+1	+2	+3	Average	t
10 Swollen glands (lymph nodes) in neck or armpit	0	9	34	283	54	15	3	0.10	2.89‡
24 Frequent or painful urination	2	11	16	305	44	19	1	0.10	2.88‡
45 Sensitivity to heat or cold	1	19	42	257	46	29	4	0.08	1.84
47 Sweating (not due to exercise)	2	16	35	263	59	20	3	0.09	2.9‡
33 Fainting spells	0	3	9	357	21	7	1	0.06	2.61‡
46 Fever or chills	0	11	35	295	39	16	2	0.05	1.42
30 Skin (including rashes)	2	16	52	253	52	19	4	0.03	0.70
32 Cuts or sores that heal slowly	3	15	38	290	44	6	2	−0.04	−1.03
Average of all symptoms	1	10	40	261	63	20	3	0.12	7.44§

Note: Positively signed changes indicate more severe symptoms reported by mail. Average column reports the mean change per person in severity. Each symptom report was coded as 0 (none), 1 (mild), 2 (moderate), or 3 (severe). Thus, the difference between mail and telephone symptom reports can range from −3 to 3. n = 398.

†Serial position of question in interview.

‡$P < 0.05$, §$P < 0.001$. Because of the numerous comparisons, P values between 0.05 and 0.001 are indeterminate with respect to statistical significance.

TABLE 2. Mean Number of Symptoms Reported

Group 1 — **By Telephone**

	None	Mild	Moderate	Severe	Sum
Sum	30.3	**6.1** $_{1a}$	**7.3** $_{2a, 2b}$	**4.3** $_{3a}$	48.0

17.7 $_{4a}$

Group 2 — **By Telephone**

By Mail	None	Mild	Moderate	Severe	Sum	
None	22.8	1.6	0.8	0.1	25.3	
Mild	4.9	2.9	1.9	0.4	**10.0** $_{1b}$	
Moderate	2.0	1.6	3.2	1.3	**8.1** $_{2b}$	**22.7** $_{4b}$
Severe	0.4	0.4	1.2	2.7	**4.6** $_{3a}$	
Sum	30.1	**6.4** $_{1a}$	**6.9** $_{2a}$	**4.6** $_{3a}$	48.0	

17.9 $_{4a}$

Note: Group 1 was interviewed only by telephone. Group 2 was interviewed by mail and then 2 weeks later by telephone. Bold numbers indicate the means included in statistical analyses. Subscripting with the same number (eg, $_1$ and $_1$) indicates 2 means that were compared, and subscripting with the same number but a *different* letter (eg, $_{1a}$ and $_{1b}$) indicates 2 means that differed from one another significantly ($P < 0.05$).

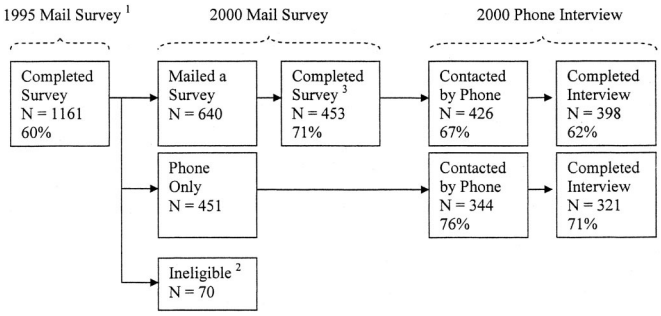

FIGURE 1. Calculation of completion rates. Ineligible participants included those who were deceased, incarcerated, hospitalized, or were not deployed to the Persian Gulf War. [1]Participants were sampled from Delaware, Illinois, New Jersey, New York, North Carolina, Ohio, and Pennsylvania. Original sample n = 2011 with n = 1935 deliverable surveys. [2]Ineligible category includes veterans who were deceased (n = 20), not deployed to the Persian Gulf War (n = 39), incarcerated (n = 3), or incapacitated (eg, due to loss of hearing) (n = 8). [3]Nonresponders to mail survey include n = 45 not contacted because of unknown address.

cation), and mostly middle-aged (mean 43 years). They reported an average of 13 moderate or severe symptoms (of 48 possible) in the interview in 1995. Veterans completing the present interviews were more likely to be white, higher rank, older, and to have initially reported more physical symptoms (t tests and χ^2, $P < .05$) than those who completed only the 1995 interview.

Participants in the present study were randomly assigned to be interviewed only by phone (completion rate 71%, n = 321), or interviewed by mail and then again 2 weeks later by phone (overall completion rate 62%, n = 398). The lower completion rates for the latter group reflect their having had 2 opportunities to decline to participate. The participants in each condition did not differ in their ethnicity, gender, socioeconomic status, rank, education, age, previous symptom reports, social desirability, functional status, or psychological distress, as examined using t tests, χ^2, and logistic regression (data shown in Appendix). Thus, random assignment successfully created equivalent experimental groups with respect to the demographic characteristics measured.

Interview Mode Varying Between Groups

Veterans reported an average of 18 symptoms by telephone and 23 by mail, $F(1,711) = 32.50$, $P < 0.001$. In support of the satisficing explanation, the effect of interview mode differed with the severity of the symptom reported. A statistically significant main effect of severity was qualified by an interaction with interview mode, $F(2, 1418) = 80.64$, 18.60, $P < .001$. Posthoc tests revealed that interview mode affected only mild symptom reports. Veterans reported an average of 6 mild symptoms by telephone and 10 by mail.

But what about social desirability? As expected, higher social desirability scores predicted fewer symptom reports of any severity, $F(1,709) = 10.11$, $P = 0.001$. However, the argument that social desirability scores explain interview-related differences in symptom reporting was not supported. Social desirability scores did not interact reliably with interview mode, and the 3-way interaction of social desirability, interview mode, and severity was also not statistically significant, $F < 1$. In other words, symptom reports made by mail and phone were equivalently related to social desirability scores. (That social desirability scores were assessed only by phone could account for *finding* a significant interaction, but it cannot account for the absence of such a finding as is reported here.) The important conclusion is that interview-related differences in symptom reporting are not explained by a propensity for socially desirable responding.

Several demographic covariates showed significant relations to symptom reporting. Older age, lower education, lower rank, and nonwhite ethnicity predicted greater symptom reporting, $F(1,709) = 6.67$ to 35.05, $P \leq .05$. Gender was unrelated to symptom reports, $F < 1$, an atypical finding that may reflect unique characteristics of women veterans. No interaction of education and interview mode was found, $F < 1$. The same findings are shown in the other analyses and are not discussed further.

Interview Mode Varying Within Subjects

The within-subjects analyses replicated the pattern of findings described above. Interviews by mail again yielded more symptom reports than those by phone, $F(1,391) = 200.07$, $P < 0.001$. The overall mean differences were nearly identical to those reported for the between-groups comparison: Veterans reported an average of 18 symptoms by phone and 23 by mail. Interview mode interacted significantly with severity, $F(2,782) = 37.90$, $P < 0.001$. As before, fewer mild symptoms were reported by phone than by mail (6 versus 10 mean symptoms). There was also a suggestion that fewer moderate symptoms were reported by phone (7 versus 8 mean symptoms).

An inspection of Table 1 shows that mail symptom reports were more likely to switch from mild to none when assessed by phone than vice versa (1.6 versus 4.9 mean symptoms). To a lesser extent, there was a similar bias towards switching from moderate to none (0.7 versus 2.0 mean symptoms). However, they did not make intermediate switches (such as going from severe to mild) as might be suggested by more diffuse socially desirable responding. The findings offer strong support for the satisficing explanation because they show that the symptoms not reported by phone were milder symptoms that are presumably harder to recall. Furthermore, there was a total failure to report some mild

symptoms rather than a general tendency to report all symptoms as less severe.

Interview mode differences were present for all 5 clusters of symptoms and for most of the individual symptoms, as shown in Table 2. Social desirability scores again predicted differences in symptom reporting, $F(1,391) = 13.96$, $P < 0.001$, and its effects were again independent of interview mode, $F \leq 1$. The relations to interview mode and social desirability scores did not differ among 5 symptom clusters ($F < 1$). In summary, the interview mode difference in symptom reporting was due to changes in the severity of symptoms reported, but not due to socially desirable responding.

Alternative Explanations

The experiment's design allows us to rule out 2 nuisance explanations for the *within*-subjects findings. In the time between the mail and phone interviews, some symptoms could have resolved. Alternatively, merely completing 1 survey could cause fewer symptoms to be later reported in any later survey. Both explanations conflict with the *between*-subjects findings. Because the phone-only group had only 1 interview, there was no previous interview and no opportunity for their symptoms to regress toward health. Thus, the groups' different levels of symptom reporting are readily interpretable as caused by interview mode. Additional disconfirming evidence against the alternative explanations would come from analyses showing that the 2 phone interviews yielded similar symptom reports. Indeed, we find that there was no main effect of having been recently interviewed about symptoms, $F<1$, and recent interview status did not interact with severity or social desirability, $F = 0.89$ to 2.08 (NS). Thus, phone groups did not differ but the phone and mail groups always differed. The findings lead us to reject the resolving symptoms and the order effect explanations.

Predictive Utility of Symptom Reports

The data show that people elide mild symptom reports in phone surveys. This caused us to wonder whether symptom reports that differ in severity are equally important in understanding a person's well being. To examine this question, we predicted physical functional status in a hierarchical, blockwise multiple regression. Demographic variables were entered into the model in the first step, followed by severe symptoms, then moderate symptoms, and finally mild symptoms. The regression was conducted 3 times, once for mail survey reports, and once each for the 2 phone surveys. Table 3 shows that severe symptoms explained 30% to 33% of the variance in functional status. Moderate symptoms explained an additional 6% to 12% of the variance. Mild symptoms explained an additional 1% to 3% of the variance. Repeating the regressions with mild symptoms entered before the other symptoms found that mild symptom reports explained about half as much variance (0.5% to 2%). It appears that, before mild symptoms become fully interpretable, one needs to know about severe symptoms.

DISCUSSION

Previous research has documented greater symptom reporting by mail than by phone, prompting speculation that socially desirable responding might explain the finding. The present study examined the social desirability explanation as well as an alternative explanation that phone surveys exaggerate respondents' tendency to satisfice—to offer what they feel is a "good enough" response that may be well short of a best or most complete answer. One of the strengths of the study was our between- and within-subjects replication of the interview mode effect.

We replicated the well-established finding that fewer symptoms are reported in phone than in mail interviews and that those with higher trait social desirability scores are less likely to report symptoms. However, the 2 findings were unconnected to one another. Socially desirable responding did *not* explain the difference in symptom reporting related to interview mode. Instead, we found that the effect was explained by an interaction with symptom severity. Veterans were less likely to report mild symptoms by phone than by mail and were somewhat less likely to report moderate symptoms as well.

TABLE 3. Usefulness of Symptom Reports in Predicting Functional Status

	Group 1	Group 2	
	Phone R^2 Change	Mail R^2 Change	Phone R^2 Change
Step 1: demographic covariates	7%*	8%*	8%*
Step 2: severe symptoms	32%*	30%*	34%*
Step 3: moderate symptoms	11%*	7%*	10%*
Step 4: mild symptoms	3%*	2%*	1%*

Note: The dependent variable was functional status as measured by the physical component summary of the SF-36.
*$P < 0.001$

The pattern of results is consistent with the satisficing hypothesis—that responding over the phone adds cognitive demands that make satisficing more likely.[29] One possibility is that phone interviewers regulate the pace of the interview, communicating a need for parsimonious responses and also limiting the time that participants have to recall the relevant information and formulate an answer. There are economic motivations for the companies that conduct phone interviews to train their staff to complete calls quickly because faster calls translate into higher productivity and profits. It is also possible that the nature of the interview itself constrains respondents by forcing them to hold in mind both the question asked and the response options. Although greater cognitive sophistication might limit such an effect, we did not find that the effect of interview mode varied with the veterans' education levels. Lastly, the findings are also consistent with participants recalling but not reporting some symptoms as a way to shorten the interview.

The present study adds to our understanding of the role of social desirability in symptom reports, both as an item and a person trait. We replicated the previously documented finding that persons with higher trait social desirability scores are less likely to report health problems (Brewer NT, Hallman WK, Fiedler N, et al., submitted for publication) but the effect was independent of interview mode. It is noteworthy that potentially embarrassing symptoms (eg, incontinence, diarrhea, sexual problems, and painful urination) did not appear to show a stronger interview mode effect. This observation offers additional support for the conclusion that socially desirable responding (whether trait- or question-based) is not a good explanation for interview mode effects in symptom reporting.

The present findings suggest a strategy for assessing medical symptoms through self-report. First, dismissing, or discounting, reports of mild symptoms may be useful. The interview mode bias was primarily in mild symptom reports, and they predicted minimal additional variance in physical function. Second, adjusting for the effects of social desirability as we have defined it here may be more complex. Researchers have argued that social desirability scores measure a substantive personality characteristic and statistically "controlling" for its influence may mean eliminating a substantive portion of a measure of health.[23,25,35] Additional research is needed showing the conditions under which social desirability influences responses to better understand when and how to intervene.

The findings of the present study are subject to several limitations. Interview mode was manipulated experimentally and thus its effect on mild symptom reporting can be interpreted as causal, but such an inference would be strengthened by a replication of the experimental findings. The social desirability findings are cross-sectional and should be viewed with some skepticism regarding causality and the possibility of a third causal variable such as the personality trait neuroticism. Second, we do not have a gold standard of accurate symptom reporting and so cannot corroborate the satisficing account that phone reports of mild symptoms are biased. Future research could employ a more direct measure of cognitive effort to clarify the finding. Third, we interviewed military veterans with unexplained medical symptoms who may be unique in unanticipated ways. Gulf War veterans report more symptoms than comparison groups, and veterans on the Registry probably overrepresent those with higher symptom reports. Soldiers may also be more homogenous than the general population because the military selects recruits who are uniformly healthy through rigorous medical screening and homogenizes them psychologically through extensive training. The generalizability of the present findings to randomly sampled military populations and relatively healthy populations has yet to be established. Fourth, the study examines social desirability scores assessed by phone. There may be different findings for socially desirability scores assessed by mail. More importantly, variations in the *difference* between social desirability scores obtained by mail and phone could potentially explain the interview mode effect. Given the small effect sizes associated with social desirability scores, we doubt that this is the case, but the possibility merits future research. Last, the shortened version of the social desirability scale that we used had only modest internal consistency and this may have reduced our ability to find interactions with it.

In summary, 3 variables showed clear moderating effects on symptom reporting: method of interview, social desirability, and symptom severity. Veterans reported fewer symptoms in phone than mail interviews. They were less likely to report additional symptoms if they showed higher social desirability scores, but this effect was independent of interview mode. The interview mode effect was mainly the result of fewer mild symptom reports. The findings place social desirability as an independent but small factor in symptom reporting and support the satisficing hypothesis.

ACKNOWLEDGMENTS

The authors thank Daniel Wartenburg, Kendal Boyd, Srinivas Maloor, and Gozde Ozakinci for their work on the study and Erich Labouvie for his help with the statistics.

REFERENCES

1. Brewer NT, Chapman GB, Brownlee S, et al. Cholesterol, adherence, and illness cognition. *Br J Health Psychol*. 2002;7:433–447.
2. Meyer D, Leventhal H, Gutmann M. Common-sense models of illness: the example of hypertension. *Health Psychol*. 1985;4:115–135.
3. Fukuda K, Nisenbaum R, Stewart G, et al. Chronic Multisymptom illness affecting Air Force veterans of the Gulf War. *JAMA*. 1998; 280(11):981–988.
4. Hodgson MJ, Kipen HM. Gulf War illnesses: causation and treatment. *J Occup Environ Med*. 1999;41:443–452.
5. Sigal L, Hassett AL. Contributions of societal and geographical envi-

ronments to "chronic Lyme disease": a psychopathogenisis and aporology of a new "medically explained symptoms" syndrome. *Environ Health Perspect*. 2002;100(Suppl 4):607–611.
6. Wyer RS, Srull TK. *Memory and Cognition and Its Social Context*. Hillsdale, NJ: Lawrence Erlbaum; 1989.
7. Tourangeau R, Rips L, Rasinki K. *The Psychology of Survey Response*. New York: Cambridge University Press; 2000.
8. McHorney CA, Kosinski CM, Ware JE. Comparison of the costs and quality of norms for the SF-36 Health Survey collected by mail versus telephone interview: results from a national survey. *Med Care*. 1994; 32:551–567.
9. Siemiatycki J. A comparison of mail, telephone, and home interview strategies for household health surveys. *Am J Public Health*. 1979;69: 238–245.
10. Moum T. Mode of administration and interviewer effects in self-reported symptoms of anxiety and depression. *Social Indicators Res*. 1998;45: 279–318.
11. Parry HJ, Crossley HM. Validity of responses to survey questions. *Public Opinion Quarterly*. 1950;14:80.
12. Locander WB, Sudman S, Bradburn N. An investigation of interview method, threat, and response distortion. *J Am Stat Assoc*. 1976;71:269–275.
13. Aquilino WS, Loscuito LA. Interview mode effects in drug use surveys. *Public Opinion Quarterly* 1990;362:395.
14. Hochstim JR. A critical comparison of three strategies of collecting data from households. *J Am Stat Assoc*. 1962;62:976–989.
15. Paulhus DL. Two-component models of socially desirable responding. *J Personality Social Psychol*. 1984;43:598–609.
16. Paulhus DL. Measurement and control of response bias. In: Robinson JP, Shaver PR, Wrightsman LS, editors. *Measures of Personality and Social Psychological Attitudes*. San Diego: Academic Press; 1991:17–59.
17. Crowne DP, Marlowe DA. A new scale of social desirability independent of psychopathology. *J Consulting Psychol*. 1960;24:349–354.
18. Bardwell W, Ancoli-Isreal S, Dimsdale JE. Response bias influences mental health symptom reporting inpatients with obstructive sleep apnea. *Ann Behav Med*. 2001;12:313–317.
19. Carnrike CLM. Is social desirability associated with symptom distress in organ transplant candidates? *Psychol Health Med*. 1997;2:243–250.
20. Carstensen LL, Cone JD. Social desirability and the measurement of psychological well-being in elderly persons. *J Gerontol*. 1983;38:713–715.
21. Klassen D, Hornstra RK, Anderson PB. Influence of social desirability on symptom and mood reporting in a community survey. *J Consulting Clin Psychol*. 1975;43:448–452.
22. Koller M, Heitman K, Krussman J, Lorenz W. Symptom reporting in cancer patients II: Relation to social desirability, negative affect, and self-reported health behaviors. *Cancer*. 1999;86:1609–1620.
23. McCrae RR, Costa PT. Social desirability scales: more substance than style. *J Consulting Clin Psychol*. 1983;51:882–888.
24. Taylor SE, Lerner JS, Sherman DK, et al. Portrait of a self-enhancer: Well adjusted and well liked or maladjusted and friendless? *J Personality Social Psychol*. 2003;84:165–176.
25. Taylor SE, Brown JD. Illusion and well-being: a social psychological perspective on mental health. *Psychol Bull*. 1988;103:193–210.
26. Richman W, Weisband S, Kiesler S, et al. A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. *J Appl Psychol*. 1999;84:754–775.
27. Holtgraves T. Social desirability and self-reports: testing models of socially desirable responding. *Personality Social Psychol Bull*. 2004;30:161–172.
28. Simon HA. *Models of Man*. New York: Wiley; 1957.
29. Krosnick JA. Response strategies for coping with the cognitive demands of attitude measures in surveys. *Appl Cognitive Psychol*. 1991;5:213–236.
30. Higgins ET, Rholes WS, Jones CR. Category accessibility and impression formation. *J Exp Social Psychol*. 1977;13:141–154.
31. Hallman WK, Kipen HM, Diefenbach M, et al. Symptom patterns among Gulf War Registry veterans. *Am J Public Health*. 2003;93:624–630.
32. Reynolds WM. Development of reliable and valid short forms of the Marlowe-Crowne Social Desirability Scale. *J Clin Psychol*. 1982;38:119–125.
33. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). *Med Care*. 1992;30:473–483.
34. Derogatis LR, Spencer PM. *BSI Administration and Procedural Manual I*. Baltimore: Clinical Psychometric Research; 1982.
35. Kozma A, Stones MJ. Social desirability in measures of subjective well-being: a systematic evaluation. *J Gerontol*. 1987;42:56–59.

Appendix Demographic Characteristics of Samples

	Completed 2000 Interview		Condition	
	No (n = 372)	**Yes (n = 719)**	**Phone Only (n = 321)**	**Mail-Phone (n = 398)**
% white ethnicity	66%	81%†	81%	81%
% male gender	90%	92%	91%	94%
Military rank (% enlisted)	94%	88%	89%	87%
Years education	–*	12.3	12.2	12.3
Age	40.4	42.5†	42.2	42.7
Medical symptoms	15.8	12.7†	12.6	12.7
Marlowe–Crowne Score	–*	6.0	6.0	6.0
Physical function	–*	43.5	43.5	43.6
Psychological distress	–*	0.57	0.55	0.58

Note: Medical symptoms reflect the number of moderate or severe symptoms of 48 assessed that were reported in a previous interview in 1995. Physical function assessed using the SF-36. Psychological distress assessed using the Brief Symptom Inventory.
*Not assessed in 1995 interview.
†$P < 0.05$

