HEALTH

THE ARTS

CHILD POLICY

CIVIL JUSTICE

EDUCATION

ENERGY AND ENVIRONMENT

HEALTH AND HEALTH CARE

INTERNATIONAL AFFAIRS

NATIONAL SECURITY

POPULATION AND AGING

PUBLIC SAFETY

SCIENCE AND TECHNOLOGY

SUBSTANCE ABUSE

TERRORISM AND
HOMELAND SECURITY

TRANSPORTATION AND
INFRASTRUCTURE

WORKFORCE AND WORKPLACE

This PDF document was made available from www.rand.org as a public service of the RAND Corporation.

[Jump down to document ▼](#)

The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world.

Support RAND

[Browse Books & Publications](#)

[Make a charitable contribution](#)

For More Information

Visit RAND at www.rand.org

Explore [RAND Health](#)

View [document details](#)

This product is part of the RAND Corporation reprint series. RAND reprints reproduce previously published journal articles and book chapters with the permission of the publisher. RAND reprints have been formally reviewed in accordance with the publisher's editorial policy.

The medical group survey from the CAHPS® (formerly Consumer Assessment of Health Plans Study) project, G-CAHPS, focuses on patient experiences in receiving care from their medical group practice. We compared mail and telephone responses to the G-CAHPS survey in a sample of 880 patients from four physician groups. Patients were randomly assigned to mode. Analyses included comparison of response rates, missing data, internal consistency reliability of six multi-item scales, and mean scores. A total of 537 phone completes and 343 mail completes were obtained (54% response rate). There were no significant differences in internal consistency by mode. In addition, there was only one significant mode difference in item and composite means by mode of administration after adjusting for case-mix differences. This study indicates that mail and telephone modes of data collection for the G-CAHPS survey produce similar results.

Keywords: *mode effects; CAHPS; mail administration; telephone administration*

COMPARISON OF MAIL AND TELEPHONE IN ASSESSING PATIENT EXPERIENCES IN RECEIVING CARE FROM MEDICAL GROUP PRACTICES

KIMBERLY A. HEPNER
*RAND Corporation,
Health Sciences Program*

JULIE A. BROWN
RAND Corporation, Survey Research Group

RON D. HAYS
*University of California Los Angeles School
of Medicine, and RAND Corporation,
Health Sciences Program*

AUTHORS' NOTE: This article was supported by grant number 5 U18 HS-00924 from the Agency for Healthcare Research and Quality. Kimberly A. Hepner, Ph.D., and Ron D. Hays, Ph.D., were also supported, in part, by the UCLA/DREW Project EXPORT, National Institutes of Health, National Center on Minority Health & Health Disparities, (P20-MD00148-01). Ron D. Hays, Ph.D., was also supported, in part, by UCLA Center for Health Improvement in Minority Elders/Resource Centers for Minority Aging Research, National Institutes of Health, National Institute of Aging, (AG-02-004). Send correspondence concerning this article and reprint requests to Kimberly A. Hepner, RAND Corporation, 1776 Main Street, Santa Monica, CA 90407; e-mail: hepner@rand.org.

Patients' reports and ratings of their health care experiences have become an integral part of evaluating the quality of health care. Accrediting agencies (e.g., National Committee for Quality Assurance, Joint Commission on Accreditation of Healthcare Organizations) now require health plans and hospitals to measure patient perspectives when evaluating quality of services provided. Furthermore, employers are routinely provided with patient assessments of care to assist them in selecting health plans for their employees. In recent years, there have also been increased efforts to provide summaries of patient-reported measures of quality and satisfaction directly to consumers in an effort to inform decisions in selecting a health plan, provider group, or individual provider.

Because of the increased emphasis on patient-reported experiences in evaluating the quality of health services, it is important to use reliable and valid survey measures. One potential source of error variance is variance attributable to the method of data collection. The most common methods are mail and telephone. It has also been suggested that a mix of these methodologies, such as mail with telephone follow-up, may lead to increased response rates and provide a better representation of the sampled population (Fowler et al., 2002). Combining data collected from differing methods (either within a single study or when evaluating findings across studies) assumes that responses from these methods are comparable. Although efforts can be made to construct an instrument with the goal that it will produce comparable results using different data collection methodologies, it is essential to evaluate this presupposition.

Mail and telephone administration have different advantages and disadvantages. Telephone administration is typically associated with increased response rates and decreased rates of missing data (Burroughs, Waterman, Cira, Desikan, & Dunagan, 2001; Harris, Weinberger, & Tierney, 1997; Perkins & Sanson-Fisher, 1998), while mail administration is typically associated with decreased cost (McHorney, Kosinski, & Ware, 1994; Walker & Restuccia, 1984). Comparisons of the reliability of responses have not consistently favored telephone or mail (McHorney et al., 1994; Perkins & Sanson-Fisher, 1998).

There is some evidence that survey responses vary by mode of data collection. When patients are asked to report on their experiences in receiving care, their satisfaction with care, or their current health status, responses tend to be more favorable to a phone interviewer

compared to self-administration (Burroughs et al., 2001; Fowler, Roman, & Di, 1998; McHorney et al., 1994; Walker & Restuccia, 1984). This has been attributed to a greater demand for socially desirable responses when responding to an interviewer as opposed to self-administering an instrument (Dillman, 2000). Social desirability pressures may be particularly salient in response to sensitive items (e.g., Turner, Lessler, & Devore, 1992). However, a comparison of telephone and mail administration of the CAHPS® (formerly the Consumer Assessment of Health Plans Study) 1.0 health plan study, Fowler, Gallagher, & Nederend (1999) found generally comparable results across modes.

This article compares mail and telephone responses to the Group-Level CAHPS (G-CAHPS) survey. The G-CAHPS survey focuses on patient experiences in receiving care from their medical group practice and has not been previously evaluated for comparability across methods of administration. We compare the two modes of administration in terms of response rates, missing data, internal consistency of multi-item scales (composites), and reports and rating of care.

METHOD

A random sample of 1,641 patients were selected across four physician groups (approximately 400 patients per group) in California: two multispecialty group practices, one safety net provider, and one independent practice association (IPA) that included multiple physician practices. Sample size was chosen based on a desire to obtain 350 completes per mode. With 350 cases per mode, an alpha of .05, two-tailed test, we would have 80% power to detect a small effect (.20; Cohen, 1988). Sampling frames were constructed from practice management systems and included patients at least age 18 years with at least one visit to the doctor's office in a 12-month period. The sampling frame from one of the four physician groups did not include telephone numbers. Telephone number look-up was conducted using various methods including automated phone look-up vendors, directory assistance, and Internet-based databases. All participants received a U.S.\$10 payment for participation. The payment was mailed after completion of the survey. Participants were surveyed between July and October 2000.

Patients were randomly assigned to one of two administration modes: mail or telephone. Mail-mode participants received an initial mailing, a reminder postcard 2 weeks later, and a second mailing of the questionnaire to nonresponders approximately 4 weeks after the reminder postcard. Telephone interviews were attempted with patients assigned to mail administration if they did not complete the instrument by mail 3 weeks after the third and final mailing. The data collection protocol specified a maximum of five telephone attempts per case during follow-up with mail nonresponders.

Data collection was conducted in English and Spanish. Sampling information did not include information on preferred language (English or Spanish). However, one physician group, the safety net provider, was known to serve a large number of patients who would require a Spanish-language version of the survey. Mailings to patients sampled from this group included materials in English and Spanish. For the three remaining physician groups, little was known about the proportion of patients who would require Spanish-language materials. Mailings to patients sampled from these groups included a pre-addressed, postage-paid Spanish-language postcard patients could return to request materials in Spanish.

Telephone-mode participants received an advance letter followed 1 week later by an attempt to conduct a telephone interview. The advance letter was printed in English on one side and in Spanish on the other side. For cases assigned to telephone as primary data collection mode, the data collection protocol specified a maximum of 10 attempts per case. The protocol for all telephone data collection required that attempts must be made across several time slots (week-day morning, afternoon and evening, weekends) before a case could be finalized as a noncontact or not completed case. Telephone interviewers conducted interviews reading from a questionnaire in the language (English or Spanish) identified during the informed consent script. The decision to use a paper questionnaire rather than a computer-assisted interview was based on cost.

G-CAHPS

This study used a version of the CAHPS survey that is designed to assess patients' experiences with medical provider groups (G-CAHPS). The G-CAHPS instrument is similar to the CAHPS 2.0 by

design. While the G-CAHPS survey is similar to the CAHPS 2.0 core survey, the G-CAHPS survey focused on domains that patients indicated were important in evaluating medical groups. Development of the G-CAHPS included review of existing measures, focus groups, cognitive interviews, and field testing (Solomon, Hays, Zaslavsky, & Cleary, 2005). The G-CAHPS assesses six multi-item composites to elicit reports about care in the last 12 months: (a) wait times (4 items); (b) access to needed care (2 items); (c) communication between doctors and patients (4 items); (d) coordination between primary care physicians and specialists (2 items); (e) courtesy and respect shown by the office staff (2 items); and (f) advice on preventive health (2 items). Response options for these items are yes or no, never/sometimes/usually/always, or a big problem/a small problem/not a problem. In addition, the survey included three global rating items assessing the personal doctor or nurse, specialist, and all care in the last 12 months. These items ask respondents to select a number from 0 (*worst care possible*) to 10 (*best care possible*). Spanish language instruments were produced using the CAHPS translation approach described by Weidmer, Brown, and Garcia (1999). Evaluation of the psychometric properties of the G-CAHPS instrument provided support for its reliability and validity in measuring medical group performance (Solomon et al., 2005). Additional information about the CAHPS family of surveys is available online (www.cahps-sun.org).

ANALYSIS PLAN

Parallel analyses were conducted to evaluate the presence of mode effects. First, we present a comparison of mail versus phone. These analyses compared patients who responded using the mode to which they were randomly assigned. Specifically, patients randomized to the mail mode who responded via the mail mode were compared with patients randomized to the phone mode who responded via the phone mode. Because mail nonresponders were eventually contacted by phone, these analyses excluded patients who were initially randomized to the mail mode that responded to a telephone follow-up (mixed-mode phone respondents). We then present a secondary comparison of randomized to phone versus mixed-mode phone. Specifically, patients who were randomized to phone were compared with patients randomized to mail but responded through phone follow-up.

The G-CAHPS includes screeners so that respondents answer only items that are applicable to them. Analyses of missing data focused on inappropriate missing—items that a respondent should have answered based on their responses to screener items. The proportion of inappropriate missing was computed for each respondent. The denominator included the items that should have been answered by the respondent, and the numerator was the number of those items that were missing. Missing data by mode was compared using *t* tests. As noted above, telephone interviews were not administered via a computer-assisted instrument, which introduced the possibility of missing data because of interviewer error.

Cronbach's alphas for each of the six composite scales were compared by mode (Feldt, Woodruff, & Salih, 1987). Significance levels were adjusted for multiple comparisons using the Hochberg (1998) stepped procedure for an initial alpha level of .05.

Differences in item and composite means by mode were evaluated using *t* statistics, adjusting for multiple comparisons. We also evaluated mode differences in item and composite means after adjusting for respondent characteristics in regression models. To evaluate differences in respondent characteristics across mode and their potential utility as case-mix variables, nine respondent characteristics (age, gender, language, race/ethnicity, education, income, overall ratings of health, quality of life, and health care in the United States) were compared by mode using *t* tests (continuous variables by mode), chi-square tests (categorical variables by mode), or correlations (with items and composites), as appropriate. All nine respondent characteristic variables were significantly different ($p < .05$) between modes in either the mail versus phone comparison or the randomized to phone versus mixed-mode phone comparison, or were significantly related ($p < .05$) to at least one item or composite. Therefore, all nine respondent characteristics were included in the case-mix models.

RESULTS

A total of 777 completes were obtained through the mode to which they were randomized (47% response rate; 343 to mail, 434 to phone). Including those study participants who were randomized to mail mode first but completed a telephone interview increased the number

of phone completes to 537 and overall sample size to 880 (54% response rate). Of the 820 patients assigned to the telephone mode, 53% ($n = 434$) completed an interview. Of the 821 patients assigned to mail mode, 42% ($n = 343$) returned a completed mail survey, and an additional 13% (103) completed a telephone interview, resulting in an overall response rate of 55% for this group.

Rates of active refusal were low and comparable across the two sample groups. Six percent of the patients assigned to telephone mode and 5% of the patients assigned to mail as the primary mode actively refused to participate in the study (i.e., returned a blank mail survey, contacted study staff to request removal from the study mailing list, or declined participation through a phone interviewer). One difference by mode is the identification of patients as deceased, as speaking a language other than English or Spanish, or being otherwise ineligible to complete a survey (11% for telephone vs. .5% for mail). The identification of patients who cannot be reached via the assigned mode is another area in which mail and telephone differed. Of the patients assigned to the telephone mode, 20% could not be reached by telephone (including nonworking numbers, unpublished numbers, and wrong numbers) whereas surveys were undeliverable to 7% of the patients assigned to mail as the primary mode.

Response rates across the four physician groups ranged from 34% to 61%. Lower response rates were obtained from the safety net provider (34%) and the site that was not able to provide telephone numbers (36%). Response rates of 59% and 61% were obtained from the remaining two groups.

MAIL VERSUS PHONE

Patients randomly assigned to mail or telephone provided similar proportions of inappropriately missing data, Mail $M = .049$, Phone $M = .040$, $t(742) = 1.25$, $p = .205$. After adjusting for multiple comparisons, there were no significant differences in internal consistency reliability by mode for any of the six composites. Cronbach's alpha ranged from .56 to .87 for mail respondents and from .53 to .87 for phone respondents.

Table 1 provides the unadjusted mean and standard deviation for each item and composite for patients randomized to mail versus phone. After adjusting for multiple comparisons, one significant dif-

TABLE 1
Mail Versus Phone:
Unadjusted Item, Composite, and Global Rating
Means and Standard Deviations by Mode

<i>Item</i>	<i>Mail</i>	<i>Mail N</i>	<i>Phone</i>	<i>Phone N</i>
Access: Getting care quickly	3.06 (.71)	312	3.09 (.74)	393
Q6 Care as soon as wanted	3.36 (.80)	208	3.35 (.89)	231
Q8 Get care for regular care	3.18 (.89)	267	3.23 (.90)	320
Q9 Wait more than 15 minutes	2.67 (1.01)	269	2.68 (1.13)	315
Q28 Get advice needed	3.05 (.90)	206	3.11 (.96)	254
Access: Getting needed care	2.66 (.58)	314	2.67 (.57)	408
Q13 Problem to get care	2.67 (.63)	304	2.66 (.62)	400
Q45 Problem to get referral	2.61 (.69)	214	2.61 (.68)	238
Communication	3.43 (.59)	310	3.46 (.68)	401
Q14 Doctors listen	3.44 (.72)	308	3.50 (.79)	401
Q16 Doctors explain	3.57 (.62)	307	3.57 (.73)	398
Q21 Show respect	3.46 (.70)	304	3.49 (.81)	398
Q22 Spend enough time	3.26 (.79)	308	3.27 (.88)	397
Coordination	2.97 (.97)	207	3.01 (.98)	220
Q50 Help decide specialist	3.02 (1.20)	206	2.87 (1.26)	218
Q51 Informed	2.91 (1.10)	199	3.17 (1.02)	207
Office Staff	3.39 (.69)	329	3.41 (.77)	416
Q29 Office staff respect	3.51 (.67)	329	3.51 (.78)	415
Q30 Staff helpful	3.26 (.79)	327	3.32 (.86)	415
Preventive counseling	1.45 (.45)	283	1.45 (.44)	326
Q39 Foods you eat	1.42 (.49)	265	1.49 (.50)	294
Q40 Exercise	1.51 (.50)	282	1.58 (.49)	324
Global rating items				
Q52 Specialist rating	8.14 (2.26)	203	8.39 (2.16)	229
Q42 Personal doctor or nurse rating	8.52 (1.89) ^a	281	8.95 (1.53) ^a	332
Q25 All care rating	8.06 (1.94)	305	8.21 (1.95)	396

NOTE: a. Mean is significantly different across mail and phone after adjusting for multiple comparisons.

ference between mail and phone remained. Specifically, global rating of personal doctor or nurse was significantly higher for phone respondents, $t(611) = 3.13, p = .002$. This would be considered a small effect (Cohen, 1988).

Of the potential case-mix adjustment variables, mode differences were observed for self-reported overall quality of life, rating of health care in the United States, and Hispanic ethnicity in the mail versus phone comparison. Each of these was higher in the patients randomized to phone. The randomized to phone versus mixed-mode phone comparison also demonstrated differences in language of survey

(English vs. Spanish) and Other race (a combination of low-frequency races including Blacks, Native Americans, Native Hawaiians, and Pacific Islanders). Furthermore, all of the nine respondent characteristics were significantly related to at least one of the items or composites. Therefore, all nine respondent characteristics were included in the case-mix model. After case-mix adjustment and adjustment for multiple comparisons, only one significant association of mode with reports or ratings of care remained. Specifically, the effect of mode on global rating of personal doctor or nurse was significant in the mail versus phone comparison, $t(571) = -3.28$, $p = .001$. This would be considered a small effect (Cohen, 1988).

RANDOMIZED TO PHONE VERSUS MIXED-MODE PHONE

Respondents randomized to phone yielded a higher proportion of inappropriately missing data, randomized to phone $M = .040$, mixed-mode phone $M = .024$, $t(493) = 2.20$, $p = .028$. This is a very small effect.

After adjusting for multiple comparisons, there were no significant differences in internal consistency reliability by mode for any of the six composites. Cronbach's alpha ranged from .53 to .87 for patients randomized to phone and from .62 to .82 for mixed-mode phone respondents.

Table 2 provides the unadjusted mean and standard deviation for each item and composite by mode. There were no significant differences for items or composites by mode; therefore, adjustment for multiple comparisons was unnecessary. After adjustment for age, gender, language, race/ethnicity, education, income, overall ratings of health, quality of life, and health care in the United States, and adjusting for multiple comparisons, there were no significant associations of mode with reports or ratings of care.

DISCUSSION

These results suggest that the G-CAHPS survey provides generally comparable results by telephone and mail. Rates of missing data and estimates of internal consistency reliability were indistinguishable by mode for both methods of comparison. All 16 report items, all six

TABLE 2
Randomized to Phone Versus Mixed-Mode Phone:
Unadjusted Item, Composite and Global Rating
Means and Standard Deviations by Mode

<i>Item</i>	<i>Phone Random</i>	<i>Phone Random N</i>	<i>Mixed Phone Respond</i>	<i>Mixed Phone N</i>
Access: Getting care quickly	3.09 (.74)	393	3.09 (.77)	82
Q6 Care as soon as wanted	3.35 (.89)	231	3.29 (.86)	51
Q8 Get care for regular care	3.23 (.90)	320	3.23 (.86)	65
Q9 Wait more than 15 minutes	2.68 (1.13)	315	2.61 (1.17)	67
Q28 Get advice needed	3.11 (.96)	254	3.13 (.93)	52
Access: Getting needed care	2.67 (.57)	408	2.67 (.59)	93
Q13 Problem to get care	2.66 (.62)	400	2.63 (.64)	92
Q45 Problem to get referral	2.61 (.68)	238	2.65 (.66)	49
Communication	3.46 (.68)	401	3.44 (.64)	92
Q14 Doctors listen	3.50 (.79)	401	3.38 (.91)	92
Q16 Doctors explain	3.57 (.73)	398	3.55 (.70)	92
Q21 Show respect	3.49 (.81)	398	3.49 (.75)	92
Q22 Spend enough time	3.27 (.88)	397	3.34 (.82)	92
Coordination	3.01 (.98)	220	3.06 (1.01)	53
Q50 Help decide specialist	2.87 (1.26)	218	2.98 (1.22)	53
Q51 Informed	3.17 (1.02)	207	3.13 (1.16)	53
Office staff	3.41 (.77)	416	3.43 (.74)	101
Q29 Office staff respect	3.51 (.78)	415	3.62 (.74)	100
Q30 Staff helpful	3.32 (.86)	415	3.26 (.89)	101
Preventive counseling	1.45 (.44)	326	1.56 (.46)	79
Q39 Foods you eat	1.49 (.50)	294	1.54 (.50)	69
Q40 Exercise	1.58 (.49)	324	1.64 (.48)	78
Global rating items				
Q52 Specialist rating	8.39 (2.16)	229	8.58 (2.32)	55
Q42 Personal doctor or nurse rating	8.95 (1.53)	332	8.94 (1.55)	78
Q25 All care rating	8.21 (1.95)	396	8.04 (2.14)	92

composites, and two of three global ratings did not differ by mode of administration. One difference was observed when comparing respondents randomized to mail with respondents randomized to phone (mail vs. phone comparison). Patients provided higher ratings of their "personal doctor or nurse" when responding to a telephone interview. The difference in the global doctor rating remained significant even after adjusting for case mix. No differences were found when comparing patients randomized to phone with those who responded to phone after not responding to mail. It is important to note that the current

study results may not generalize to newer variants of telephone data collection such as interactive voice response (IVR).

Previous research has indicated that the tendency to provide more socially desirable, or positive, responses is stronger when responding to a telephone interviewer compared to a self-administered mail survey (Burroughs et al., 2001; Fowler, 1998; McHorney, 1994; Walker & Restuccia, 1984). Only one item in the current study, however, demonstrated this tendency. Previous research has demonstrated that patients are more likely to provide positive responses when asked about their own medical care than when asked to rate the care received by others, and that this difference is attributable to a socially desirable response set (Hays & Ware, 1986). The global rating of personal doctor or nurse is the only item that refers to the provider as "your personal doctor or nurse." Other items do not include this personal referent (e.g., referring only to doctors and other health care providers). Further evaluation of alternate wording would be useful to determine whether the inclusion of a specific personal referent increases socially desirable responding in patients reporting on their health care experiences.

We conducted supplementary analyses to compare mail respondents to all phone respondents (randomized to phone and mixed phone respondents combined) and found similar results to the primary mail versus phone comparison results presented here. After adjusting for multiple comparisons, global ratings of personal doctor or nurse were significantly higher for phone respondents. After adjusting for case-mix differences, the difference in the global doctor rating remained significant. Fortunately, in these evaluations of the G-CAHPS, differences by mode were not present for the majority of items and composites after adjusting for important case-mix variables.

The comparability of instruments administered by different modes is important, in part, because it provides greater flexibility to survey users to select data collection methodologies based on their unique situation. For example, telephone mode may be most appropriate for low-literacy populations or when the survey user does not have access to accurate address information to support a mail survey; however, telephone administration may be associated with higher costs. Mixed mode may be used to improve response rate and mitigate some costs by first relying on mail administration before commencing with telephone follow-up. Survey users with few resources may opt for mail-

only administration. Our results provide support for the comparability of these methods.

Limited availability of accurate information on preferred language continues to be a problem faced by survey administrators and researchers. Our approach was the most cost-effective method available to deal with the fact that we had no information regarding language preference, other than the general language profile of the group practice. We only provided double-sided (English/Spanish) surveys to one of the group practices known to have a high proportion of Spanish-speaking patients. Postcards to request a Spanish-language questionnaire were included with English questionnaires sent to patients from the other three group practices. An alternative methodology would have been to send double-sided questionnaires to the entire sample. We decided against this because the cost associated with this methodology was too high. We believe the best solution to this problem is to encourage group practices to collect accurate language preference information from their patients.

When self-report instruments are used to assess patient experiences in receiving care, it is essential to evaluate whether responses vary by mode of administration. Comparability across mode of administration is particularly important because mail, telephone, and mixed modes (e.g., mail with telephone follow-up) are all used commonly. The current study demonstrates that the mail and telephone modes of data collection yield similar results for the G-CAHPS. Hence, it is possible to use both modes of data collection when comparing patient evaluations of care provided by different physician groups. One item, the global doctor rating, may differ by administration mode and caution should be used when comparing responses to this item across modes.

REFERENCES

- Burroughs, T. E., Waterman, B. M., Cira, J. C., Desikan, R., & Dunagan, W. C. (2001). Patient satisfaction measurement strategies: A comparison of phone and mail methods. *Journal on Quality Improvement*, 27(7), 349-361.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Dillman, D. A. (2000). *Mail and Internet surveys* (2nd ed.). New York: John Wiley.

- Feldt, L. S., Woodruff, K. J., & Salih, F. A. (1987). Statistical inference for coefficient alpha. *Applied Psychological Measurement, 11*, 93-103.
- Fowler, F. J., Gallagher, P. M., & Nederend, S. (1999). Comparing telephone and mail responses to the CAHPS survey instrument. *Medical Care, 37*(3), MS41-MS49.
- Fowler, J. F., Gallagher, P. M., Stringfellow, V. L., Zaslavsky, A. M., Thompson, J. W., & Cleary, P. D. (2002). Using telephone interviews to reduce nonresponse bias to mail surveys of health plan members. *Medical Care, 40*(3), 190-200.
- Fowler, F. J., Roman, A. M., & Di, Z. X. (1998). Mode effects in a survey of Medicare prostate surgery patients. *Public Opinion Quarterly, 62*(1), 29-46.
- Harris, L. E., Weinberger, M., & Tierney, W. M. (1997). Assessing inner-city patients' hospital experiences: A controlled trial of telephone interviews versus mailed surveys. *Medical Care, 1*, 70-76.
- Hays, R. D., & Ware, J. E. (1986). My medical care is better than yours. *Medical Care, 24*(6), 519-524.
- Hochberg, Y. (1998). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika, 75*, 800-802.
- McHorney, C. A., Kosinski, M., & Ware, J. E. (1994). Comparisons of the costs and quality of norms for the SF-36 Health Survey collected by mail versus telephone interview: Results from a national study. *Medical Care, 32*(6), 551-567.
- Perkins, J. J., & Sanson-Fisher, R. W. (1998). An examination of self- and telephone-administered modes of administration for the Australian SF-36. *Journal of Clinical Epidemiology, 51*(11), 969-973.
- Solomon, L., Hays, R. D., Zaslavsky, A., & Cleary, P. D. (2005). Psychometric properties of the Group-Level Consumer Assessment of Health Plans Study (G-CAHPS) instrument. *Medical Care, 43*(1), 53-60.
- Turner, C. F., Lessler, J. T., & Devore, J. (1992). Effects of mode of administration and wording on reporting of drug use. In C. F. Turner, J. T. Lessler, & J. C. Gfoerer (Eds.), *Survey measurement of drug use: Methodological studies* (pp. 177-220). Washington, DC: Government Printing Office.
- Walker, A. H., & Restuccia, J. D. (1984). Obtaining information on patient satisfaction with hospital care: Mail versus telephone. *Health Services Research, 19*(3), 291-306.
- Weidmer, B., Brown, J., & Garcia, L. (1999). Translating the CAHPS 1.0 survey instruments into Spanish. *Medical Care, 37*(3), MS89-MS96.