Published in final edited form as:

Psychol Assess. 2001 June ; 13(2): 216–229.

Confirmatory Factor Analysis of the Consumer Assessment of Health Plans Study (CAHPS) 1.0 Core Survey

Grant N. Marshall

RAND, Health Sciences

Leo S. Morales

University of California at Los Angeles, and RAND, Health Sciences

Marc Elliott

RAND, Health Sciences

Karen Spritzer

University of California at Los Angeles

Ron D. Hays

University of California at Los Angeles, and RAND, Health Sciences

Abstract

The National Consumer Assessment of Health Plans Study (CAHPS) Benchmarking Database was used to assess the factor structure and invariance of the CAHPS 1.0 Core Survey. Separate analyses were conducted with Latino and non-Latino Caucasian consumers drawn from commercial and Medicaid sectors ($N = 15,092$). Results demonstrated that the 23 CAHPS 1.0 report items measure consumer reports of experiences with 5 aspects of health plan performance: Access to Care, Timeliness of Care, Provider Communication, Health Plan Consumer Service, and Office Staff Helpfulness. Four items assessed global ratings of care. Analyses revealed an identical pattern of fixed and free factor loadings across all samples. Magnitude of factor loadings and correlations among factors was essentially equivalent within a common health service sector. A higher order factor analysis revealed that rating and reports of care showed marked convergence.

A confluence of interrelated social forces, including the rise of consumerism, the advent of managed care, the emergence of the government as a primary payer of health services, and an increasingly competitive health care marketplace has spurred interest in consumer evaluations of health services. Consumer judgments about health care have emerged as an essential index

Correspondence concerning this article should be addressed to Grant N. Marshall, RAND, 1700 Main Street, P.O. Box 2138, Santa Monica, California 90407-2138. Electronic mail may be sent to Grant_Marshall@rand.org..

Grant N. Marshall and Marc Elliott, RAND, Health Sciences, Santa Monica, California; Leo S. Morales and Ron D. Hays, RAND, Health Sciences, Santa Monica, California, and Division of General Internal Medicine and Health Sciences Research, Department of Medicine, University of California at Los Angeles; Karen Spritzer, Division of General Internal Medicine, Department of Medicine, University of California at Los Angeles.

The Consumer Assessment of Health Plans Study (CAHPS) is funded by the U.S. Agency for Healthcare Research and Quality (AHRQ) and the Health Care Financing Administration through cooperative agreements with Harvard Medical School, RAND (U18HS09204), and the Research Triangle Institute. Manuscript preparation was supported, in part, by a Minority Investigator Supplement funded by the AHRQ. The comparative CAHPS survey data used in this analysis were provided by the National CAHPS Benchmarking Database (NCBD), a collaborative initiative of the Quality Measurement Advisory Service, The Picker Institute, and Westat. The NCBD is funded by the AHRQ under Contract 290-95-2005. Additional information about the study can be obtained by calling the AHRQ Clearinghouse at 800-358-9295 or by accessing the following website: <http://www.ahrq.gov/qual/cahpsix.htm/>. A copy of the entire CAHPS 1.0 Adult Core Survey has been reproduced in Appendix I of *Medical Care*, 37, MS106-MS109, 1999. We thank project officers Chris Crofton, Chuck Darby, and Beth Kosiak for their active participation and thoughtful feedback throughout the project. We also thank Paul Cleary for comments and suggestions on an earlier version of this article.

of health care quality and a significant determinant of marketplace viability. Although health behavior is determined by multiple factors, consumer evaluations have proved at least modestly predictive of various health-related behaviors, including the initiation of malpractice litigation (Penchansky & MacNee, 1994; Vaccarino, 1977), poor adherence to medical regimens (Sherbourne, Hays, Ordway, DiMatteo, & Kravitz, 1992), and disenrollment from health plans and providers (Newcomer, Preston, & Harrington, 1996; Rossiter, Langwell, Wan, & Rivnyak, 1989; Schlesinger, Druss, & Thomas, 1999). For example, using a randomly selected nationwide survey sample of approximately 3,000 elderly fee-for-service and health maintenance organization (HMO) Medicare beneficiaries, Rossiter et al. reported that about 50% of all plan disenrollment was primarily attributable to consumer dissatisfaction. Perhaps not surprisingly, consumer evaluations have been implicated as both an antecedent and consequence of good health (e.g., Hall, Roter, & Milburn, 1999; Marshall, Hays, & Mazel, 1996).

Research interest in patient satisfaction dates at least as far back as the 1940s and 1950s when sociologists actively began to examine this construct in the context of physician-patient relations (Locker & Dunt, 1978). Nonetheless, the systematic development of psychometrically sound measures of patient satisfaction is a relatively recent phenomenon. Although ad hoc measures of satisfaction have been and continue to be created, one of the earliest rigorous programs of research on the measurement of patient satisfaction with health care is exemplified by the development of the Patient Satisfaction Questionnaire (PSQ). Developed and refined by Ware and colleagues over a period of years (Ware, Snyder, & Wright, 1976a, 1976b), the initial measure consisted of 80 items and was intended to be applicable in general population studies and to be useful for planning, administration, and evaluation of health services delivery programs (Ware et al., 1983). Continued revision of the PSQ resulted in various versions used in national surveys (Aday, Andersen, & Fleming, 1980; Aday, Fleming, & Andersen, 1984) as well as the RAND Health Insurance Experiment (Davies, Ware, Brook, Peterson, & Newhouse, 1986). A third-generation instrument, the PSQ-III, consisting of 50 items, remains in relatively widespread use (e.g., Hueston, Mainous, & Schilling, 1996; Marshall et al., 1996).

In all of its various permutations, the PSQ consisted of a series of statements reflecting beliefs about key dimensions of satisfaction with doctors and health care services (e.g., technical and interpersonal skills of providers, waiting time for appointments, office waits, emergency care, costs of care, insurance coverage, availability of hospitals, and other resources) and satisfaction with care in general. Items were general, rather than specific, in nature (e.g., doctors are good about explaining the reason for medical tests), with respondents indicating the degree to which they agreed or disagreed with each item. Responses were provided without regard to a specific time frame and without reference to a specific physician or health care setting. Moreover, the focus of the PSQ was not on evaluation of specific health care experiences but on global appraisals of care. Although the PSQ played a seminal role in the development of research assessing patient satisfaction with health care, continued advances in the conceptualization and measurement of consumer evaluations of health care have shifted the measurement focus in a different direction.

The Consumer Assessment of Health Plans Study (CAHPS; Cleary & Edgman-Levitan, 1997; Hays et al., 1999), which constitutes the next generation of instruments designed to assess consumer evaluations of health care, is currently the most widely used instrument. Adopted by the National Committee for Quality Assurance as part of its accreditation process (National Committee for Quality Assurance, 2001), CAHPS data were available to assist the decision making of an estimated 90 million Americans in 1999 (Cleary, 1999). The CAHPS 1.0 Core Survey possesses several desirable features. First, the CAHPS 1.0 assesses both reports about specific health care experiences (e.g., delays in gaining access to care) and specific health care

providers and settings. CAHPS 1.0 also retains the ability to measure global ratings of care received (e.g., overall judgments concerning one's health plan). Incorporation of reports of health care experiences can help to target specific domains in need of quality improvement, thereby addressing the criticism that general ratings are too nebulous to yield practically meaningful information (e.g., Williams, 1994). By including items about both specific experiences and global impressions, it is possible to examine empirically the association between the two types of information. In addition, although many preceding instruments have required respondents to provide answers without regard to a specific time frame (e.g., Kippen, Strasser, & Joshi, 1997; Marshall, Hays, Sherbourne, & Wells, 1993), the CAHPS 1.0 items have been expressly anchored to a specific reference period (i.e., the past 6 months). Given the interpretative difficulties inherent in attempting to evaluate change over time when items are answered without respect to a specific time frame, incorporation of a temporal reference frame offers clear advantages (Cleary et al., 1998). The relative brevity of the CAHPS 1.0 survey also constitutes a significant advantage over previous instruments such as the PSQ.

Preliminary studies have indicated that the CAHPS 1.0 survey instrument possesses promise as a tool for measuring consumer appraisals of health care (e.g., Hays et al., 1999; Zaslavsky, Beaulieu, Landon, & Cleary, 2000). For example, Hays et al. provided basic information about the psychometric properties of the CAHPS 1.0, using samples of nearly 6,000 Medicaid recipients and approximately 11,000 persons with private health insurance. Internal consistency reliability estimates for various indexes were generally acceptable, with median estimates of 0.70 and 0.76 in the Medicaid and private health insurance groups. Attesting to the construct validity of the CAHPS 1.0, global health care ratings were significantly correlated with consumers' willingness to recommend the plan to family and friends ($r = 0.75, p < .01$) and to re-enroll in the plan if given the choice ($r = 0.77, p < .01$), in a subset of 313 respondents receiving health care through Medicaid.

Despite the apparent promise of the CAHPS 1.0, key questions have yet to be answered. First, fundamental assumptions underlying the development of the instrument have not yet been demonstrated. In particular, although the CAHPS 1.0 was explicitly designed to assess consumer experiences with respect to multiple dimensions of care, there has, as yet, been no evaluation of the factor structure of the instrument at the level of the individual.¹ Thus, the multifactorial structure of the CAHPS 1.0 has been assumed but not demonstrated.

Moreover, although supplemental items have been developed to tailor the CAHPS to the assessment needs of various vulnerable populations, including persons on Medicaid or Medicare and children (Brown, Nederend, Hays, Short, & Farley, 1999; Schnaier et al., 1999; Shaul et al., 1999), the core set of CAHPS 1.0 items was intended to measure the same constructs across diverse populations. As yet, this assumption of measurement equivalence across diverse groups has not been tested explicitly. Thus, demonstration of the equivalence of the core instrument across persons from different ethnic groups would provide much needed information concerning the validity of the instrument.

An additional issue concerns the degree of concordance between alternative methods of summarizing experiences with health care (Marshall et al., 1993). The CAHPS 1.0 instrument incorporates two strategies. The CAHPS 1.0 includes items that assess global perceptions of care by asking consumers to indicate the extent to which they agree with statements reflecting beliefs about broad domains of care (e.g., overall rating of health care provider). These items

¹We note that a recent exploratory factor analysis of the CAHPS 1.0 has examined the associations among dimensions using health plans, rather than individuals, as the unit of analysis (Zaslavsky et al., 2000). In other words, the responses of individuals receiving care within a given health plan were averaged, and the resulting plan level ratings were factor analyzed. Although this approach addresses important research questions, the resulting factor structure would not necessarily correspond to that which would emerge from an individual level analysis.

are intended to be aggregated, yielding an index of global ratings of health care. The CAHPS 1.0 also incorporates an additional strategy that calls for consumers to provide reports about the frequency with which specific experiences took place within multiple dimensions of care. In principle, each of these strategies should have different utilities, depending on the goals of the assessment. In particular, the first measurement strategy may be useful for providing global consumer perceptions, while requiring comparatively few items and thus minimizing respondent burden. In contrast, the second approach provides highly detailed information about specific content domains that may be of relevance to targeted quality improvement efforts, albeit at the expense of greater respondent burden.

Whereas the two measurement approaches (i.e., ratings vs. reports) differ with respect to both level of specificity (specific vs. global) and the dimension of measurement (i.e., frequency vs. general perceptions), the two strategies might be expected, when viewed from a broader perspective, to yield comparable information. That is, at a more abstract level of analysis, information gleaned by ratings and reports should be essentially equivalent insofar as consumer views regarding health care in general can be inferred by combining reports regarding specific health experiences to yield a single summary index. Given the trade-offs associated with these two strategies, direct comparison of the concordance of the two approaches with respect to global consumer judgments would provide valuable insight into the conceptualization and measurement of health care experiences.

The goals of this study were to confirm (a) the individual-level factor structure of the CAHPS 1.0 adult survey, (b) the invariance of the factor structure across persons from White and Latino ethnic groups as well as individuals receiving care within commercial and Medicaid health service sectors, and (c) the concordance between judgments about health care derived from consumer reports of specific health care experiences and consumers' global ratings of health care.

Although we recognize that it is important to understand the health care experiences of other ethnic minority groups (e.g., Asian-Pacific Islanders, African Americans, and indigenous Americans), our decision to focus on White (non-Latino) and Latino health care consumers was influenced by the emergence of Latinos as the largest and fastest growing minority segment of the U.S. population (Day, 1996). Moreover, although a number of studies have documented the barriers to health care faced by Latinos (Ginzberg, 1991; Phillips, Mayer, & Aday, 2000; Valdez, Giachello, Rodriquez-Trias, Gomez, & de la Rocha, 1993), few studies have attempted to evaluate experiences with the health care system encountered by Latinos. Assessment of the health care experiences of Latinos, including evaluation of the equivalence of consumer assessment surveys, is particularly important insofar as this population has unique cultural needs that may not be well served by a health care system oriented toward meeting the needs of members of the dominant culture (Molina, Zambrana, & Aguirre-Molina, 1997).

Method

Data Source

The data were drawn from adult surveys of the National CAHPS Benchmarking Database (NCBD 1.0). The purpose of the NCBD 1.0 is to provide a frame of reference against which to compare the results obtained by other CAHPS users. Health care plans and sponsors participated voluntarily in this data-pooling effort and included Medicaid as well as commercial participants. No overlap exists between data used in this study and any previously published research using the NCBD 1.0 database.

The NCBD Medicaid database included information from 29 HMOs and two primary care case management programs located in the District of Columbia, Arkansas, Kansas, Minnesota,

Oklahoma, and Washington. The NCBD database also included information from commercially insured persons drawn from 27 HMOs, eight physician provider organizations, three point-of-service plans, and one fee-for-service plan. Data from 15 additional commercial health plans, for which information needed for more detailed classification of plan type was unavailable, were obtained from the District of Columbia, Florida, Kansas, New Jersey, Oklahoma, and Washington. Program participation in the NCBD 1.0 was voluntary, and information concerning the number or characteristics of health plans that chose not to participate is not available.

Data were collected, using telephone interviews and mail surveys, from 28,354 enrolled health service consumers between 1997 and 1998. Limitations in the NCBD CAHPS 1.0 data do not allow us to determine the number of respondents completing interviews versus mail surveys. Thus, response rates by the two modes of data collection are unavailable. Preliminary research does suggest, however, that CAHPS information obtained via these two methods are substantially similar. In particular, Fowler, Gallagher, and Nederend (1999) reported the results of three preliminary studies involving approximately 1,300 respondents drawn from the Medicaid and commercial sectors, indicating that only 2 of the 27 items used in the current study showed mode-of-administration effects. Moreover, the direction of these effects differed across the three studies. Finally, surveys were administered in either English or Spanish, as necessary. The NCBD data did not contain information, however, as to the language version used by a given respondent.

Participants

Participants completing the survey did so voluntarily and were guaranteed of the confidentiality of their responses. The overall response rate for the survey was 52%, with rates of 66% and 34% for commercial and Medicaid participants, respectively. For the purposes of the present study, participants included Latino and White (non-Latino Caucasian) enrollees who were at least 18 years of age. Respondents were assigned to racial-ethnic categories based on their answers to the following questions: “Are you of Hispanic or Spanish family background?” (yes, no), and “How would you describe your race?” (American Indian or Alaskan Native, Asian or Pacific Islander, Black or African American, White, Another race or multiracial). Respondents who answered yes to the first item were categorized as Latino, regardless of race. Respondents who answered no to the first item and White to the latter item were categorized as non-Latino Caucasian.

Four separate analytic samples were created: Latinos drawn from the commercially insured ($n = 1,020$) and Medicaid sectors ($n = 609$) and whites drawn from the commercially insured ($n = 7,983$)² and Medicaid sectors ($n = 5,480$). Respondents were classified as receiving care within either commercially insured or Medicaid sectors on the basis of administrative data from health plans and sponsors. Age and gender for each of the four analytic samples are shown in Table 1. As anticipated, Medicaid participants were somewhat younger and disproportionately female (Health Care Financing Administration, 1999a, 1999b).

Measures

Embedded in a larger health survey consisting of 57 questions, the CAHPS 1.0 Adult Core Survey includes 23 items measuring reported experiences concerning specific aspects of health plan performance, derived from a review of the literature, focus groups, cognitive interviews, and field tests (Harris-Kojetin, Fowler, Brown, Schnaier, & Sweeny, 1999). Twenty-one of the 23 report items are assessed using a 4-point scale, with response options ranging from 1

²This sample was randomly drawn from a total sample of 14,643 cases. Current versions of EQS for Windows (Bentler, 1995) can only accommodate sample sizes of fewer than 8,000 cases.

(*never*) to 4 (*always*). Two additional report items are answered using a 2-point (yes = 1, no = 2) response format. The instrument also contains four items assessing global ratings of health care (i.e., health plan, quality of care, personal doctor, and specialists). Items assessing ratings differ from those measuring reports in that the former refer to global assessments rather than judgments about specific health care experiences. Each of the four rating items is answered using an 11-point scale, with anchor points reflecting either the worst (0) or the best (10) possible care. Question wording and response options are shown in the Appendix. As shown in the Appendix, the four rating items were answered without regard to a specific time frame, whereas the 23 reporting items were answered with respect to the previous 6 months. The decision to use a specific frame of reference was made to facilitate measurement of change over time. The 6-month time frame was adopted to minimize recall problems that might occur from asking respondents to remember experiences over a longer time period (e.g., 12 months). A more complete discussion of the issues of interest in choosing to adopt a 6-month time frame can be found in Harris-Kojetin et al. (1999).

On the basis of instrument development work (McGee et al., 1999) and psychometric analysis of the CAHPS 1.0 (Hays et al., 1999), we hypothesized that the 23 reporting items about specific aspects of care represented five underlying dimensions: Access to Care (five items), Timeliness of Care (five items), Provider Communication (six items), Health Plan Consumer Service (five items), and Office Staff Helpfulness (two items). Descriptive statistics for individual items for each of the four analytic samples, organized by a priori scale membership, are shown in Tables 2 and 3. These statistics reflect item wording as found in the Appendix. For purposes of factor analysis, however, all items were recoded so that higher scores signify more positive evaluations.

Data Analysis

Missing data and their treatment—In analyzing the CAHPS 1.0 Survey findings, two types of missing data were encountered. One type of missing data could be considered to be missing in the “traditional” sense. In other words, certain questions that were posed to respondents should have been—but were not—answered, a widely encountered problem in survey research. Simply put, a small number of respondents failed to complete all relevant items on the survey, and thus certain data points were inappropriately missing. A second type of information was also, in the strictest sense, unavailable, but was not missing as the term is conventionally used to refer to responses that are inappropriately missing (i.e., items that should have been answered). A more apt description of this latter type of missing data might be “appropriately missing” or missing because of inapplicability of the question to the respondent. More specifically, inclusion of a past-6-months frame of reference for most questions meant that the survey instrument was constructed in such a way that all questions were not equally relevant to all respondents. In other words, respondents had had different degrees of exposure to their health care systems in the preceding 6 months. The survey was crafted expressly to accommodate these differences in experience with the health care system. Thus, if a given respondent had not had a particular experience in the last 6 months, then they were either instructed to skip the item, or to indicate that the question was not applicable to them. All of the items appropriately afforded respondents one of these two options for not providing an answer to questions for which they had no experience on which to base a response.

Of the two sources of missing data, the vast majority of missing items were absent because they did not apply to the respondent. In contrast, only 5% of missing data were inappropriately missing. Because the rate of true missing data was equivalent across groups, small in magnitude, and within traditionally acceptable limits, it did not appear that the degree of inappropriately missing data constitutes a significant data quality problem. Ninety percent of

all data imputation was performed for appropriately— that is, not applicable—rather than inappropriately, missing items.

Although inappropriately missing data did not constitute a major problem by conventional standards, casewise deletion of all cases, including those who omitted questions that were irrelevant to them because of lack of experience with the health care system, would have resulted in a significant loss of data. Moreover, invoking casewise deletion, irrespective of its impact on sample size, would have been highly questionable, inasmuch as it would have resulted in a very unique sample of extremely high health service utilizers, a subpopulation unlikely to be representative of the full group under study. Given our goal of conducting factor analyses to draw inferences from the full sample, it was clear that some imputation strategy was required.

We considered several data imputation methods. Two commonly employed strategies, namely, mean imputation and regression imputation, were determined to be inappropriate. Mean substitution, perhaps the most common form of data imputation, is widely recognized as underestimating the magnitude of variances and covariances (Little & Rubin, 1987). With mean imputation, the variance is zero for all imputed scores. When conducting factor analysis, it is critical that one's imputation strategy not alter variances and covariances. In fact, some researchers have gone so far as to conclude that mean substitution never be used (Graham, Hofer, Donaldson, MacKinnon, & Schafer, 1997). Given that the primary aim of the present research is to estimate covariation among items, mean substitution is clearly inappropriate for our purposes. Regression imputation, another commonly used approach, was also rejected on the grounds that this strategy also artificially reduces variance, albeit to a lesser degree than mean imputation (Little & Rubin, 1987).

After considering these various options for data imputation, we selected a hot-deck imputation strategy (Rubin, 1987; for an extended discussion, see Brick & Kalton, 1996) to impute data for both appropriately and inappropriately missing responses. Although seldom seen in the psychological literature, this approach is well suited to the aims of the present study. The basic approach of the hot-decking strategy is to replace missing values with a randomly sampled value drawn from the original population. This approach preserves the variance of values and does not lead to biased parameter estimates as would mean or regression imputation. Hot-deck imputation will introduce “noise” by increasing the standard error of parameter estimates. Although increasing standard errors does not result in systematic bias, it will reduce statistical power. Given the relatively large analytic samples used here, however, loss of power does not constitute a significant problem. Finally, there is no theoretical threshold concerning the maximum number of cases for which data can be imputed using the hot-decking approach.

In this particular application of hot-deck imputation, respondents were grouped into quintiles based on the average of their responses to two survey items intended to capture global perceptions of health services (i.e., Q9, global rating of health care provider, and Q45, global rating of health plan). This stratification by global ratings of care further refined the sampling strategy by limiting the random draw of values to persons who were similar with respect to global ratings of their provider and health plan. In other words, for each missing item, a randomly selected value was drawn without replacement from among respondents in the same overall satisfaction quintile who had answered that item. The randomly selected value was then used in place of the missing item.

It should be noted in passing that our use of hot-decking imputation for appropriately missing items was not based on the assumption that data were missing at random. In fact, as noted previously, the survey instrument had been intentionally constructed so that respondents would not answer questions for which they had no experience on which to base responses. Thus, the

appropriately missing data were clearly not missing at random. Instead, the rationale underlying our imputation strategy for the vast majority of data was that participants who did not have certain health care experiences in the previous 6 months would have responded similarly, had they had those experiences, to persons with otherwise comparable overall health care ratings. Although the imputation of inappropriately missing items does presume that the data are missing at random, the magnitude of inappropriately missing data was judged to be small and equivalent for all subgroups (5%) and thus should not introduce significant bias.

Overview of analytic method—Confirmatory factor analysis of latent variables, using the EQS structural equation modeling program, constituted the principal method of data analysis (Bentler, 1995).³ Within structural equation modeling, hypotheses are translated into a series of regression equations that can be solved simultaneously to generate an estimated covariance matrix. By means of various goodness-of-fit indexes, including the normed (NFI) and non-normed fit (NNFI) indexes (Bentler & Bonett, 1980), the comparative fit index (CFI; Bentler, 1990), the incremental fit index (IFI; Bollen, 1989), and the root mean squared error of approximation (RMSEA; Browne & Cudeck, 1993), the estimated matrix can be evaluated against the observed sample covariance matrix to determine whether the hypothesized model is an acceptable representation of the data. In general, incremental fit indexes (i.e., NFI, NNFI, CFI, and IFI) above 0.90 signify good model fit. RMSEA values lower than .08 signify acceptable model fit, with values lower than .05 indicative of good model fit (Browne & Cudeck, 1993).

Analytic plan—Maximum likelihood estimation was used with raw data as input. One of the assumptions underlying maximum likelihood estimation is that the data conform to multivariate normality. Univariate skewness and kurtosis data shown in Table 3 reveal some departure from normality at the level of the individual item. This departure at the univariate level manifested itself in a relatively moderate degree of departure from multivariate normality as determined by Mardia's coefficients (Mardia, 1970; Mardia, 1974). Robust estimation conducted in preliminary single-sample models revealed that nonnormality did not have a marked effect on parameter estimates. Unfortunately, robust estimation procedures are not currently permissible within EQS for multisample models. In any event, results of several studies suggest that maximum likelihood estimation is remarkably robust even in the face of marked departure of data from multivariate normality (e.g., Huba & Harlow, 1987; Olsson, Foss, Troye, & Howell, 2000; Wang, Fan, & Willson, 1996).

To determine the extent to which a common factor structure accurately characterized White and Latino subsamples in both health care sectors, a series of confirmatory factor analyses was conducted. Our sequential approach to modeling was in general accord with the strategy described by Byrne (1994). The first phase separately modeled all four analytic samples. Two subsequent parallel series of multi-sample analyses were conducted, focusing separately on respondents receiving care within the commercial and Medicaid sectors. We chose to conduct separate analyses on the commercial and Medicaid sector patients for two reasons. First, as is well known, the clients served by these two sectors are quite different. In particular, the Medicaid population is disproportionately composed of young women with dependent children who qualify for special government programs (e.g., Aid for Families with Dependent Children). Moreover, the benefit structures and characteristics of the services provided by Medicaid and commercial sectors are themselves quite different. Inasmuch as both the clients served by the

³As noted previously, 2 of the 23 items included in our analytic data set had binary response formats. Although continuous—rather than dichotomous—data should be used in factor analysis when possible (Comrey, 1978), factor analysis of binary data is not necessarily problematic (e.g., Mislevy, 1986; Parry & McArdle, 1991). We chose to retain the two binary items so as to maintain the fidelity of the CAHPS 1.0. Selected factor analyses in which the two binary items were omitted, however, indicated that inclusion of these items did not substantially alter the pattern of findings reported here.

two sectors and the benefit structure and other characteristics of the sectors themselves differ, we had no a priori basis for positing factorial invariance of the CAHPS 1.0 across service sectors.

Our initial hypothesized representation of the pattern of fixed and free factor loadings underlying latent constructs was based on previous conceptual considerations (McGee et al., 1999) and empirical findings (Hays et al., 1999). A series of models was built, each adopting a progressively more stringent set of assumptions about measurement invariance drawn from classical measurement theory (Lord & Novick, 1968). As an initial starting point, we tested a five-factor, configurally invariant model (Horn, McArdle, & Mason, 1983) in which each item was posited to contribute to one, and only one, factor. The essential assumption of configural invariance is that the pattern of fixed and free factor loadings is constant across groups. In other words, whereas the pattern of freely estimated and fixed factor loadings was identical across groups, the estimated parameters were free to vary.

After determining that the reporting data conformed to assumptions of configural invariance, we then assessed the degree to which the data fit the requirements of so-called weak factorial invariance (Meredith, 1993). The core assumption underlying weak factorial invariance is that analogous factor loadings across multiple groups are invariant with respect to their contribution to the latent factor. In other words, not only is the pattern of fixed and freely estimated loadings the same, but also the actual values for the estimated loadings are constrained to be equal across groups. For example, the contribution of a specific item to its respective factor (e.g., the loading for item Q5 on Access to Care) is constrained to be equal across samples. The next sequence of analyses examined the degree to which factor correlations within health service sector were invariant across Latino and non-Latino White samples. As an additional step, we then tested whether the data were consistent with assumptions underlying a cross-sample extension of parallelism (Lord & Novick, 1968). Parallelism is predicated on a more stringent definition of invariance that incorporates the additional restriction that error terms for analogous items are equivalent across groups. As noted by Byrne (1994, p. 161) and others (e.g., Bentler, 1995, p. 151), this step is generally considered to be a very strict test of equivalence and is usually not done. After establishing the measurement model that provided the best fit for the report data for both the commercial and Medicaid samples, we then examined whether reports regarding specific health care experiences in various domains formed empirically, as well as conceptually, distinct constructs. At issue in the latter analysis is whether variation among the items can be explained with fewer than five latent constructs.

In an additional series of analyses, we first examined the degree to which these reports of health care experiences converge, at a higher level of analysis, to form a single index reflecting global reports about health care experiences. We then developed a measurement model for the four items assessing global ratings of care and assessed the degree to which this measurement model was stable across ethnicity and service sectors. Finally, we determined the extent to which the rating and reporting methods of measuring global evaluations yielded concordant information.

Results

Factorial Invariance

As discussed above, separate confirmatory factor analyses were initially conducted on each of the four subsamples to assess the structure of reports of health care experiences. As noted above, the 23 report items were posited to represent five factors: Access to Care, Timeliness of Care, Provider Communication, Health Plan Consumer Service, and Office Staff Helpfulness. Correlations among the five factors were freely estimated. In the initial models, all items were allowed to load freely on their hypothesized factors but were not allowed to load on other factors. The variances for each of the factors were fixed at unity to identify the model.

Each of these models fit the data well (fit indexes not shown). Factor loadings and correlation coefficients for these four models are shown in Tables 4 and 5, respectively. As shown in Table 4, with a single exception, all factor loadings were statistically significant in all four subgroups. One item (Q38) failed to contribute significantly to the Health Plan Consumer Service factor in the Latino Medicaid sample. Interestingly, this item contributed only weakly, albeit significantly, to this dimension in the remaining three analytic samples. Nonetheless, because this loading was statistically significant in three of the four subsamples, the item was retained in all four samples to facilitate comparison across groups. Close inspection of the sign of factor loadings also revealed a single item reflecting Access to Care (Q8) for which the sign differed as a function of service sector.

As an initial reference point for subsequent two-group analyses, configural invariance was assessed for Latino and White consumers within both the commercial and Medicaid sectors. As shown in Table 6, this model (Model 1) fit the data well for both the commercial and Medicaid sectors. In the next step, cross-group equality constraints were imposed on analogous factor loadings to assess the degree to which the data conformed to weak factorial invariance. As shown in Table 6 (Model 2), this model also fit the data well for both sectors, although a chi-square difference test indicated that the initial model provided a statistically more adequate account of the data in both the commercial sector, $\chi^2(23, Ns = 7,983 \text{ and } 1,020) = 40.07, p < .001$, and Medicaid sector, $\chi^2(23, Ns = 5,480 \text{ and } 609) = 40.19, p < .001$.

Post hoc inspection of the Lagrange Multiplier univariate and cumulative multivariate tests in the manner prescribed by Byrne (1994) was used to identify constrained parameters that, if freed, would lead to a significant improvement ($p < .05$) in model fit. Because univariate tests are conducted independently of other variables without taking covariation among other parameters into account, reliance on univariate tests in isolation substantially increases the risk of Type I error. Test statistics for releasing constraints indicated that 18 of 23 cross-sample equality constraints in the commercial sector were statistically reasonable, using a criterion of $p < .05$. Releasing equality constraints on the remaining five pairs of factor loadings, involving items Q5, Q8, Q25, Q34, and Q40, resulted in a model (Model 3) that fit the data as well as the initial model, $\chi^2(18, Ns = 7,893 \text{ and } 1,020) = 11.30, p > .05$. Similar post hoc inspection of the Lagrange Multiplier univariate and cumulative multivariate tests for releasing constraints in the Medicaid sector indicated that 20 of 23 equality constraints were statistically reasonable. Releasing equality constraints on the remaining three pairs of factor loadings, involving items Q28, Q29, and Q30, resulted in a model (Model 3) that fit the data as well as the initial model, $\chi^2(20, Ns = 5480 \text{ and } 609) = 16.64, p > .05$. Thus, these results indicate that reports of care were substantially similar across ethnic groups within the two health care sectors with respect to the pattern of fixed and free loadings and, in most instances, the actual magnitude of factor loadings.⁴ As shown in Table 5, the correlations between analogous factors reflecting reports about specific health care experiences appear generally to be of comparable magnitude across consumers receiving care within a given sector.⁵

As a formal test of the equivalence, the next phase of analyses focused on the cross-group invariance of the correlations among the five factors. In separate multisample analyses

⁴Additional tests were conducted to assess whether these data fit the requirements of cross-group parallelism. Imposition of equality constraints on cross-group error terms for both the commercial and Medicaid sectors revealed that the data did not satisfy these more stringent criteria.

⁵As discussed earlier, we had no a priori expectation of factorial invariance of the CAHPS 1.0 for persons receiving care in different health service sectors. Accordingly, our analyses focused on persons of different ethnicity receiving care in a common sector. Nonetheless, two-group analyses comparing persons of a given ethnicity (Whites vs. Latinos) receiving care in different sectors (commercial vs. Medicaid) were consistent with our decision to focus on within-sector comparisons. In particular, a comparison of White consumers receiving care in commercial and Medicaid sectors revealed that imposition of equality constraints on cross-group factor loadings resulted in a marked degradation in the fit of the model. Equality constraints were not statistically reasonable in 16 of 23 instances. A parallel analysis with Latino respondents pointed to similar, albeit less pronounced, cross-sector differences.

conducted for the commercial and Medicaid sectors, the five analogous cross-group factor correlations were constrained to equal one another. For example, within the Medicaid sector, the correlation between Access to Care and Timeliness of Care, although freely estimated, was constrained to equal the same value for both ethnic groups. As shown in Table 6, this model (Model 4) fit the data well for both the commercial and Medicaid samples. Close inspection of Lagrange Multiplier tests for the commercial sample revealed that 7 of 10 equality constraints on cross-group factor correlations proved statistically reasonable. The three exceptions to this pattern were for correlations involving Access to Care with Provider Communication, Timeliness of Care with Provider Communication, and Health Plan Consumer Service with Office Staff Helpfulness. When these three constraints were released, the resulting model (Model 5) was a significantly better fit to the data as determined by a chi-square difference test, $\chi^2(3, Ns = 7,893 \text{ and } 1,020) = 11.79, p < .05$. For the Medicaid sample, equality constraints were statistically reasonable for 8 of 10 factor correlations. The only exceptions to this general pattern were for correlations involving Provider Communication and Health Plan Consumer Service as well as Access to Care and Timeliness of Care. When these two constraints were released, the resulting model (Model 5) proved a significant improvement in model fit as determined by a chi-square difference test, $\chi^2(2, Ns = 5,480 \text{ and } 609) = 14.71, p < .05$. Thus, results indicate substantial equivalence in the magnitude of analogous cross-group factor correlations for both the commercial and Medicaid samples.

Factor Independence

As shown in Table 5, and consistent with much prior research on the consumer judgments about health care (e.g., Marshall et al., 1993), correlations among all five reporting dimensions in each analytic sample were quite high. This pattern is consistent with the possibility that one or more of these factors reflecting consumer reports about health care experiences are redundant. To examine whether fewer than five factors were needed to account for variation among the data, alternative factor structures were examined. Within each service sector, 10 alternative models were tested in which the correlation between a different pair of factors was fixed at 1.0—rather than freely estimated—to reflect the possibility that the two factors tap a single dimension. In all other respects, the alternative models were identical to one another. Fit indexes and chi-square difference tests revealed that the five-factor model was statistically superior to all other models.⁶ That is, although substantially correlated in some instances, the five reporting dimensions are, in the strictest sense, empirically differentiable, with each providing unique information.

Higher Order Structure of Consumer Reports of Health Care Experiences

The substantial covariation of the five reporting dimensions implies that these dimensions may converge, at a more abstract level of analysis, to form a single overarching domain reflecting overall judgments regarding health care experiences. To examine the tenability of this higher order model, two additional multisample models were evaluated using the commercial and Medicaid samples. In both models, each of the five first-order domain-specific reporting dimensions were represented as stemming from a single broad second-order factor. To fix the scale of these models, the variance of the higher order factor was fixed at unity, as was a single factor loading for each of the primary reporting dimensions.

Although chi-square difference tests indicate that the most comparable first-order model (Model 2) offered a statistically more adequate account of the data for both the commercial sample, $\chi^2(10, Ns = 7,983 \text{ and } 1,020) = 263.90, p < .001$, and Medicaid sample, $\chi^2(10, Ns = 5,480$

⁶The strategy of fixing the correlation between factors to 1.0, rather than respecifying items to load on one fewer factor, is commonly employed as a test of the number of required factors (e.g., Bollen & Hoyle, 1990) and can be shown to be mathematically equivalent to the latter approach. Complete fit information for all models is available from G. N. M.

and 609) = 354.20, $p < .001$, both of the higher order models fit the data well in an absolute sense (see Model 6, Table 6). For each of the four subgroups, higher order factor loadings, reflecting the contribution of each primary dimension to overall evaluations of care, were statistically significant at $p < .001$.⁷ It is a virtual truism that the judgment as to the meaningfulness and utility of a particular parameterization of a model (e.g., a first-order vs. a second-order model) rests on substantive as well as mere statistical considerations (Byrne, 1994, p. 119; Saris & Stronkhorst, 1984, p. 266). For the limited goal of determining the concordance between information obtained by consumer reports about specific health system experiences and global ratings of the quality of service, we employed a second-order representation of health reports. Our use of this alternative, higher order model for this analytic purpose is based on our judgment that this model is substantively meaningful despite the fact that the first-order model proved, in the strictest sense, to be statistically superior.

Concordance Between Methods of Assessing Overall Health Care

For the specific purpose of examining the correspondence between global ratings of quality of services and reports about specific experiences with the health care system, we modeled both ratings and reports simultaneously. Prior to assessing the comparability of information obtained using the two methods (i.e., global ratings and reports of specific experiences), separate multisample models of global ratings—separately testing commercial and Medicaid samples—were developed using the four items discussed above (i.e., ratings of one's personal doctor, specialists, quality of care, and health plan). That is, a single latent variable, namely, global ratings of health services quality, was presumed to underlie the covariation of ratings of one's personal doctor (Q9), specialist care (Q14), overall care (Q37), and health insurance plan (Q45). In each instance, the model was identified by fixing the variance of the factor to unity. Within each sample, the four global ratings items were allowed to load freely on a single factor reflecting overall appraisals of care. Cross-group equality constraints were imposed on analogous factor loadings to determine whether the model met the assumptions underlying weak factorial invariance (Meredith, 1993).

This model fit the data well for both the commercial sample, $\chi^2(8, Ns = 7,893 \text{ and } 1,020) = 70.38, p < .001$, NFI = .992, NNFI = .989, CFI = .993, IFI = .993, RMSEA = .029, and Medicaid sample, $\chi^2(8, Ns = 5,480 \text{ and } 609) = 88.01, p < .001$, NFI = .989, NNFI = .985, CFI = .990, IFI = .990, RMSEA = .034. As expected, all factor loadings were statistically significant. Close inspection of Lagrange Multiplier tests revealed, however, that some of the cross-group equality constraints on factor loadings were not statistically tenable. In the commercial sample, one of four equality constraints, involving item Q9, was subsequently released, as were two of the four constraints, involving items Q9 and Q45, in the Medicaid sample. After releasing these constraints, the final models proved slightly superior to the more constrained models for both the commercial sample, $\chi^2(1, Ns = 7,893 \text{ and } 1,020) = 4.40, p < .05$, and Medicaid sample, $\chi^2(2, Ns = 5,480 \text{ and } 609) = 24.24, p < .001$.

To evaluate the concordance between global ratings of health services quality and reports about specific experiences with the health care system, both constructs were simultaneously included in a single model. Global ratings of care were modeled as described in the preceding paragraph. That is, a single latent variable measuring global ratings of care was operationalized as reflecting the covariation of the four measured variables tapping overall ratings of one's personal provider, specialists, overall health care, and health insurance plan. In contrast, a single higher order factor reflecting overall reports about specific health care experiences was

⁷Higher order factor loadings for the four samples are as follows: Latino Medicaid: Access to Care (AC) = 0.87, Timeliness of Care (TC) = 0.88, Provider Communication (PC) = 0.90, Health Plan Consumer Service (HP) = 0.64, Office Staff Helpfulness (OS) = 0.85; White Medicaid: AC = 0.91, TC = 0.90, PC = 0.91, HP = 0.63, OS = 0.85; Latino Commercial: AC = 0.95, TC = 0.83, PC = 0.86, HP = 0.70, OS = 0.75; White Commercial: AC = .95, TC = 0.87, PC = 0.87, HP = 0.70, OS = 0.75.

modeled as underlying the covariation among five first-order factors reflecting specific consumer reports concerning experiences in the following dimensions: Access to Care, Timeliness of Care, Provider Communication, Health Plan Consumer Service, and Office Staff Helpfulness. Overall then, this model included seven latent variables: five first-order factors tapping reports about specific types of health care experiences (e.g., timeliness of care), a higher order report factor reflecting the convergence of judgments about health care experiences, and a final factor assessing global ratings of health care services. The correlation between the factor reflecting global ratings of quality of services and the second-order factor assessing overall reports about specific health care experiences was the essential parameter of interest.

As shown in Table 6, this model (Model 7) proved to be a good fit to the data in both the commercial and Medicaid samples. Moreover, the magnitudes of the correlations between ratings and reports of health care were quite high in all four samples (Latino-Commercially insured: $r = 0.95$, White-Commercially insured: $r = 0.98$, Latino-Medicaid: $r = 0.98$, and White-Medicaid: $r = 0.97$), attesting to the high degree of concordance of information yielded using the two methods. Once again, however, chi-square difference tests revealed that constraining the correlation between the two methods to 1.0 led to significant degradation in the fit of the models for both the commercial sample, $\chi^2(2, Ns = 7,983 \text{ and } 1,020) = 45.25, p < .001$, and Medicaid sample, $\chi^2(2, N = 5,480 \text{ and } 609) = 55.77, p < .001$. Thus, although substantially correlated, the global ratings of care and reports of specific health care experiences yield information that is, in the strictest sense, empirically distinguishable.

Discussion

This research employed multisample structural equation modeling to examine the CAHPS 1.0 Core Survey in Latino and White health care consumers receiving care in either commercial or Medicaid service sectors. The study had three objectives. The first goal was to examine the extent to which the factor structure of the instrument conformed to a priori expectations. As anticipated (McGee et al., 1999), results demonstrate that the CAHPS 1.0 Core Survey does, in fact, measure consumer reports of health experiences in multiple conceptually and empirically distinct spheres of health care: Access to Care, Timeliness of Care, Provider Communication, Health Plan Consumer Service, and Office Staff Helpfulness.

The second major aim of this study was to examine the degree to which the CAHPS 1.0 Core Survey instrument factor structure is invariant across individuals of Latino and non-Latino Caucasian ethnicity receiving care in the same health care sector. Four models of invariance were evaluated, each of which incorporated the assumptions of the preceding definition and thus were progressively more stringent: cross-sample stability in the pattern of fixed (i.e., zero) and free (i.e., nonzero) factor loadings, cross-sample equivalence in the actual magnitude of analogous factor loadings, equality of cross-group correlations among factors, and cross-sample equivalence in the magnitude of analogous error variances.

Results revealed substantial evidence of factorial invariance. Notably, within a given health care sector, the CAHPS 1.0 Core Survey showed configural stability with respect to the pattern of fixed and free factor loadings. Stated differently, the pattern of zero and nonzero factor loadings for the instrument was identical, irrespective of the ethnicity of respondents. Analyses also revealed evidence of considerable stability in the actual magnitude of analogous factor loadings across Latinos and Whites receiving care within a common health service sector. That is, the size of analogous factor loadings for Latinos and Whites receiving care within the commercial sector were, in most instances, virtually identical in magnitude. A similar finding held true for persons of different ethnicity receiving care within the Medicaid sector. Moreover, the correlations among the five identified factors showed a high degree of equivalence for Latino and White respondents within the two health service sectors. The most stringent

assumption of equivalence across analogous error variances was not, however, borne out by the data.

The third broad objective of this study was to examine the extent of concordance between two alternative methods of assessing overall consumer judgments regarding health care services. As discussed above, the CAHPS 1.0 provides two distinct methods of assessing global judgments. Global impressions can be quantified by asking consumers to rate their care on a small number of key domains (e.g., rating of health care provider) and aggregating across these ratings. Global judgments can also be arrived at by aggregating across separate reports of specific experiences with respect to circumscribed health care realms (e.g., access to care) to yield a single score reflecting overall reports of experiences. Although the two methods have counterbalancing strengths and weaknesses, these data indicate, from a practical standpoint, that the alternative approaches yield highly comparable information with respect to overall ratings. Thus, a small number of items assessing consumer ratings of health care may offer adequate precision as a brief, minimally burdensome, stand-alone measure of overall judgments regarding care. Of course, where more comprehensive assessment of specific consumer experiences is desired, reference to global ratings alone is not sufficient.

Although these results provide empirical support for key assumptions underlying the development of the CAHPS 1.0 Core Survey, the results also suggest that minor modifications and clarifications might be possible to further refine the instrument. In particular, one item reflecting consumer reports regarding burden associated with completing health plan paperwork (i.e., Q38) proved to be a rather weak marker of Health Plan Consumer Service. This item did not load significantly for the sample composed of Latinos receiving Medicaid care and contributed only modestly, albeit significantly, to this factor for the three other subsamples. The general pattern of findings suggests that this item may not be a good marker of Health Plan Consumer Service. Similarly, a second item, Q8, also emerged as a weak marker of Access to Care in all samples. This item assessed the reported frequency with which respondents in the commercial and Medicaid sectors were able to see their preferred health care practitioner. Moreover, this item contributed positively to the Access to Care construct in the commercial sector but negatively in the Medicaid sector. Although both items in question were retained for the present purposes, future research is warranted to further evaluate their utility.

In interpreting the results of this study, a few limitations are noteworthy. Additional research is needed, of course, to determine the equivalence of the CAHPS 1.0 Core Survey instrument for use with persons from other racial or ethnic groups, including African Americans, Asian Americans, and indigenous American peoples. Moreover, because participation in the NCBD was voluntary and invitations to participate were not plan-specific, there was no way to determine whether participating organizations are representative of those that knew about the NCBD but chose not to become involved. Moreover, in this study, Medicaid recipients were less likely to participate than were persons in the commercial sector, again raising generalizability issues. Furthermore, in the present study, no data were available concerning whether a given individual of Latino descent completed an English or a Spanish version of the survey. Thus, whereas these results provide evidence that the instrument performs quite similarly across Latino and White consumers, future investigation is needed to clarify possible differences in the responses of persons of Latino origin completing different language versions of the instrument. In addition, although the best available data from other research with the CAHPS suggests that mode-of-administration effects are likely to be modest, the available NCBD database did not allow examination of the impact of mode-of-administration on item responding or likelihood of participation. Finally, although factor analytic research is an essential tool in instrument development insofar as it assists in verifying the structure of an instrument and the latent constructs that it measures, future research must address the predictive

and clinical utility of the CAHPS 1.0 Survey with respect to key criterion variables. Unfortunately, the present data set did not include measures that would permit such an examination.

In conclusion, these findings indicate that the CAHPS 1.0 instrument generally performs in a similar fashion for White and Latino health care consumers receiving care within a given health care sector (i.e., commercial vs. Medicaid). As such, these data are consistent with recent research (Morales, Reise, & Hays, 2000) suggesting that valid comparisons regarding judgments about health services can be drawn across Whites and Latinos. The latter point is important insofar as some prior research has suggested that Whites and Latinos may respond differently to questions about health services (e.g., Aday, Chiu, & Andersen, 1980; Hayes & Baker, 1998).

References

- Aday, LA.; Andersen, R.; Fleming, GV. Health care in the U.S.: Equitable for whom?. Beverly Hills, CA: Sage; 1980.
- Aday LA, Chiu GY, Andersen R. Methodological issues in health care surveys of the Spanish heritage population. *American Journal of Public Health* 1980;70:367–374. [PubMed: 7361954]
- Aday, LA.; Fleming, GV.; Andersen, R. Access to medical care in the U.S.: Who has it, who doesn't. (Continuing CHAS research series No. 32). Chicago: University of Chicago Press; 1984.
- Bentler PM. Comparative fit indexes in structural models. *Psychological Bulletin* 1990;107:238–246. [PubMed: 2320703]
- Bentler, PM. EQS Structural Equations Program manual. Los Angeles: BMDP Statistical Software; 1995.
- Bentler PM, Bonett DG. Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin* 1980;88:588–606.
- Bollen KA. A new incremental fit index for general structural equation models. *Sociological Methods & Research* 1989;17:303–316.
- Bollen KA, Hoyle RH. Perceived cohesion: A conceptual and empirical examination. *Social Forces* 1990;69:479–504.
- Brick JM, Kalton G. Handling missing data in survey research. *Statistical Methods in Medical Research* 1996;5:215–238. [PubMed: 8931194]
- Brown JA, Nderend SE, Hays RD, Short PF, Farley DO. Special issues in assessing care of Medicaid recipients. *Medical Care* 1999;37:MS79–MS88. [PubMed: 10098562]
- Browne MW, Cudeck R. Alternative ways of assessing model fit. *Sociological Methods & Research* 1993;21:230–239.
- Byrne, BM. Structural equation modeling with EQS and EQS/Windows: Basic concepts, applications, and programming. Thousand Oaks, CA: Sage; 1994.
- Cleary PD. The increasing importance of patient surveys. *British Medical Journal* 1999;319:720–721. [PubMed: 10487981]
- Cleary PD, Edgman-Levitan S. Health care quality: Incorporating consumer perspectives. *Journal of the American Medical Association* 1997;278:1608–1612. [PubMed: 9370508]
- Cleary PD, Lubalin J, Hays RD, Short PF, Edgman-Levitan S, Sheridan S. Debating survey approaches. *Health Affairs* 1998;17:265–266. [PubMed: 9455038]
- Comrey AL. Common methodological problems in factor analytic studies. *Journal of Consulting and Clinical Psychology* 1978;46:648–659.
- Davies AR, Ware JE, Brook RH, Peterson JR, Newhouse JP. Consumer acceptance of prepaid and fee-for-service medical care: Results from a randomized controlled trial. *Health Services Research* 1986;21:429–452. [PubMed: 3759474]
- Day, JC. Population projections of the United States by age, sex, races and Hispanic origin: 1995 to 2050. Washington, DC: U.S. Census Bureau; 1996. Series P25, Population estimates and projections, Rep. No. P25–1130

- Fowler FJ, Gallagher PM, Nederend S. Comparing telephone and mail responses to the CAHPS survey instrument. *Medical Care* 1999;37:MS41–MS49. [PubMed: 10098558]
- Ginzberg E. Access to health care for Hispanics. *Journal of the American Medical Association* 1991;265:238–241. [PubMed: 1984154]
- Graham, JW.; Hofer, SM.; Donaldson, SI.; MacKinnon, DP.; Schafer, JL. The science of prevention: Methodological advances from alcohol and substance abuse research. Bryant, KJ.; Windle, M.; West, SG., editors. Washington, DC: American Psychological Association; 1997. p. 325–366.
- Hall JA, Roter DL, Milburn MA. Illness and satisfaction with medical care. *Current Directions in Psychological Science* 1999;8:96–99.
- Harris-Kojetin LD, Fowler FJ, Brown JA, Schnaier JA, Sweeny SF. The use of cognitive testing for developing and evaluating CAHPS 1.0 Core survey items. *Medical Care* 1999;37:MS 10–MS21.
- Hayes RP, Baker DW. Methodological problems in comparing English-speaking and Spanish-speaking patients' satisfaction with interpersonal aspects of care. *Medical Care* 1998;36:230–236. [PubMed: 9475476]
- Hays RD, Shaul JA, Williams VSL, Lubalin JS, Harris-Kojetin LD, Sweeny SF, Cleary PD. Psychometric properties of the CAHPS 1.0 Survey Measure. *Medical Care* 1999;37:MS22–MS31. [PubMed: 10098556]
- Health Care Financing Administration. Medicaid recipients and vendor payments by age. Washington, DC: Department of Health and Human Services; 1999a. Table 6, HCFA-2082 Report: Center for Medicaid and State Operations
- Health Care Financing Administration. Medicaid recipients as a percentage of population by sex. Washington, DC: Department of Health and Human Services; 1999b. (Table 10, HCFA-2082 Report: Center for Medicaid and State Operations)
- Horn JL, McArdle JJ, Mason R. When is invariance not invariant?: A practical scientist's look at the ethereal concept of factor invariance. *Southern Psychologist* 1983;4:179–188.
- Huba GJ, Harlow LL. Robust structural equation models: Implications for developmental psychology. *Child Development* 1987;58:147–166.
- Hueston WJ, Mainous AG, Schilling RJ. Patients with personality disorders: Functional status, health care utilization, and satisfaction with care. *Family Practice* 1996;42:54–60.
- Kippen LS, Strasser S, Joshi M. Improving the quality of the NCQA (National Committee for Quality Assurance) Annual Member Health Care Survey Version 1.0. *American Journal of Managed Care* 1997;3:719–730. [PubMed: 10169534]
- Little, RJA.; Rubin, DB. Statistical analysis with missing data. New York: Wiley; 1987.
- Locker D, Dunt D. Theoretical and methodological issues in sociological studies of consumer satisfaction with medical care. *Social Science & Medicine* 1978;12:283–292.
- Lord, FM.; Novick, MR. Statistical theories of mental test scores. Reading, MA: Addison-Wesley; 1968.
- Mardia KV. Measures of multivariate skewness and kurtosis with applications. *Biometrika* 1970;57:519–530.
- Mardia KV. Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhya*, B 1974;36:115–128.
- Marshall GN, Hays RD, Mazel R. Health status and satisfaction with medical care: Results from the medical outcomes study. *Journal of Consulting and Clinical Psychology* 1996;64:380–390. [PubMed: 8871422]
- Marshall GN, Hays RD, Sherbourne CD, Wells KB. The structure of patient satisfaction with outpatient medical care. *Psychological Assessment* 1993;5:477–483.
- McGee J, Kanouse DE, Sofaer S, Hargraves JL, Hoy E, Kleimann S. Making survey results easy to report to consumers: How reporting needs guided survey design in CAHPSTM. *Medical Care* 1999;37:MS32–MS40. [PubMed: 10098557]
- Meredith W. Measurement invariance, factor analysis, and factorial invariance. *Psychometrika* 1993;58:525–543.
- Mislevy RJ. Recent developments in the factor analysis of categorical variables. *Journal of Educational Statistics* 1986;11:3–31.

- Molina, CW.; Zambrana, RE.; Aguirre-Molina, M. The influence of culture, class, and environment on health care. In: Molina, CW.; Aguirre-Molina, M., editors. *Latino health in the U.S.: A growing challenge*. Washington, DC: American Public Health Association; 1997. p. 23-43.
- Morales LS, Reise SP, Hays RD. Evaluating the equivalence of health care ratings by whites and Hispanics. *Medical Care* 2000;38:517–527. [PubMed: 10800978]
- National Committee for Quality Assurance. Specifications for survey measures. 3. Washington, DC: NCQA; 2001. HEDIS—The Health Plan Employer Information Data Set.
- Newcomer R, Preston S, Harrington C. Health plan satisfaction and risk of disenrollment among social/HMO and fee-for-service recipients. *Inquiry* 1996;33:144–154. [PubMed: 8675278]
- Olsson UH, Foss T, Troye SV, Howell RD. The performance of ML, GLS, and WLS estimation in structural equation modeling under conditions of misspecification and nonnormality. *Structural Equation Modeling* 2000;7:557–595.
- Parry CD, McArdle JJ. An applied comparison of methods for least-squares factor analysis of dichotomous variables. *Applied Psychological Measurement* 1991;15:35–46.
- Penchansky R, MacNee C. Initiation of medical malpractice suits: A conceptualization and test. *Medical Care* 1994;32:813–821. [PubMed: 8057697]
- Phillips KA, Mayer ML, Aday LA. Barriers to care among racial/ethnic groups under managed care. *Health Affairs* 2000;19:65–75. [PubMed: 10916961]
- Rossiter LF, Langwell K, Wan TT, Rivnyak M. Patient satisfaction among elderly enrollees and disenrollees in Medicare health maintenance organizations: Results from the National Medicare Competition Evaluation. *Journal of the American Medical Association* 1989;262:57–63. [PubMed: 2659836]
- Rubin, DB. *Multiple imputation for nonresponse in surveys*. New York: Wiley; 1987.
- Saris, W.; Stronkhorst, H. *Causal modelling in nonexperimental research: An introduction to the LISREL approach*. Amsterdam: Sociometric Research Foundation; 1984.
- Schlesinger M, Druss B, Thomas T. No exit? The effect of health status on dissatisfaction and disenrollment for health plans. *Health Services Research* 1999;34:547–576. [PubMed: 10357290]
- Schnaier JA, Sweeny SF, Williams VSL, Kosiak B, Lubalin JS, Hays RD, Harris-Kojetin LD. Special issues addressed in the CAHPS Survey of Medicare managed care beneficiaries. *Medical Care* 1999;37:MS69–MS78. [PubMed: 10098561]
- Shaul JA, Fowler FJ, Zaslavsky AM, Homer CJ, Gallagher PM, Cleary PD. The impact of parents reporting about their own and their children's experiences with health insurance plans. *Medical Care* 1999;37:MS59–MS68. [PubMed: 10098560]
- Sherbourne CD, Hays RD, Ordway L, DiMatteo MR, Kravitz RL. Antecedents of adherence to medical recommendations: Results from the medical outcomes study. *Journal of Behavioral Medicine* 1992;15:447–468. [PubMed: 1447757]
- Vaccarino JM. Malpractice: The problem in perspective. *Journal of the American Medical Association* 1977;238:861–863. [PubMed: 577973]
- Valdez RB, Giachello A, Rodriquez-Trias H, Gomez P, de la Rocha C. Improving access to health care in Latino communities. *Public Health Reports* 1993;108:534–539. [PubMed: 8416110]
- Wang L, Fan X, Willson VL. Effects of nonnormal data on parameter estimates and fit indices for a model with latent and manifest variables: An empirical study. *Structural Equation Modeling* 1996;3:228–247.
- Ware, JE.; Snyder, MK.; Wright, WR. Development and validation of scales to measure patient satisfaction with medical care services, Vol 1. Part A: Review of the literature, overview of methods, and results regarding construction of scales. Springfield, VA: National Technical Information Service; 1976a. (NTIS No. PB 288–329)
- Ware, JE.; Snyder, MK.; Wright, WR. Development and validation of scales to measure patient satisfaction with medical care services. Vol.1. Part B: Results regarding scales constructed from the patient satisfaction questionnaire and measures of other health care perceptions. Springfield, VA: National Technical Information Service; 1976b. (NTIS No. PB 288–330)
- Ware JE, Snyder MK, Wright WR, Davies AR. Defining and measuring patient satisfaction with medical care. *Evaluation and Program Planning* 1983;6:247–263. [PubMed: 10267253]

Williams B. Patient satisfaction: A valid construct? *Social Science & Medicine* 1994;38:509–516. [PubMed: 8184314]

Zaslavsky AM, Beaulieu ND, Landon BE, Cleary PD. Dimensions of consumer-assessed quality of Medicare managed care health plans. *Medical Care* 2000;38:162–174. [PubMed: 10659690]

Appendix CAHPS 1.0 Survey Items and Response Options

Access to Care

Q5 With the choices your health insurance plan gives you, was it easy to find a personal doctor or nurse you are happy with? (Yes = 1; No = 2; I haven't tried to find a personal doctor or nurse for myself = NA)

Q8 In the last 6 months, how often did you have to see someone else when you wanted to see your personal doctor or nurse? (Never = 1; Sometimes = 2; Usually = 3; Always = 4; I don't have a personal doctor or nurse or I haven't tried to see my personal doctor or nurse in the last 6 months = NA)

Q11 In the last 6 months, how often did you see a specialist when you thought you needed one? (Never = 1; Sometimes = 2; Usually = 3; Always = 4; I didn't need to see a specialist in the last 6 months = NA)

Q13 In the last 6 months, was it always easy to get a referral when you needed one? (Never = 1; Sometimes = 2; Usually = 3; Always = 4; I didn't need to see a specialist in the last 6 months or I didn't need a referral in the last 6 months = NA)

Q36 In the last 6 months, how often did you get the tests or treatment you thought you needed? (Never = 1; Sometimes = 2; Usually = 3; Always = 4; I haven't needed tests or treatment in the last 6 months = NA)

Timeliness of Care

Q16 In the last 6 months, how often did you get the medical help or advice you needed when you phoned the doctor's office or clinic during the day Monday through Friday? (Never = 1; Sometimes = 2; Usually = 3; Always = 4; I haven't phoned for medical help or advice during the day Monday to Friday in the last 6 months = NA)

Q17 In the last 6 months, how often did you get that help during the day Monday through Friday without a long wait? (Never = 1; Sometimes = 2; Usually = 3; Always = 4; Not Applicable = NA)

Q19 In the last 6 months, when you tried to be seen for an illness or injury, how often did you see a doctor or other health professional as soon as you wanted? (Never = 1; Sometimes = 2; Usually = 3; Always = 4; I haven't tried to get care right away in the last 6 months = NA)

Q21 In the last 6 months, when you needed regular or routine health care, how often did you get an appointment as soon as you wanted? (Never = 1; Sometimes = 2; Usually = 3; Always = 4; I haven't tried to get an appointment for regular or routine care in the last 6 months = NA)

Q25 In the last 6 months, how often did you wait in the doctor's office or clinic more than 30 minutes past your appointment time to see the person you went to see (Never = 1; Sometimes = 2; Usually = 3; Always = 4; I haven't had any appointments in the last 6 months = NA)

Provider Communication

Q28 In the last 6 months, how often did doctors or other health professionals listen carefully to you? (Never = 1; Sometimes = 2; Usually = 3; Always = 4; I had no visits in the last 6 months = NA)

Q29 In the last 6 months, how often did doctors or other health professionals explain things in a way you could understand? (Never = 1; Sometimes = 2; Usually = 3; Always = 4; I had no visits in the last 6 months = NA)

Q30 In the last 6 months, how often did doctors or other health professionals show respect for what you had to say? (Never = 1; Sometimes = 2; Usually = 3; Always = 4; I had no visits in the last 6 months = NA)

Q31 In the last 6 months, how often did doctors or other health professionals spend enough time with you? (Never = 1; Sometimes = 2; Usually = 3; Always = 4; I had no visits in the last 6 months = NA)

Q32 In the last 6 months, how often did doctors or other health professionals know what you thought they should know about your medical history? (Never = 1; Sometimes = 2; Usually = 3; Always = 4; I had no visits in the last 6 months = NA)

Q34 In the last 6 months, how often were you involved as much as you wanted in these decisions about your health? (Never = 1; Sometimes = 2; Usually = 3; Always = 4; No decisions were made about my health care in the last 6 months = NA)

Health Plan Consumer Service

Q38 In the last 6 months, how often did you have more forms to fill out for your health insurance plan than you thought was reasonable? (Never = 1; Sometimes = 2; Usually = 3; Always = 4; I had no visits = NA)

Q40 In the last 6 months, how often did your health insurance plan deal with approvals or payments without taking a lot of your time and energy? (Never = 1; Sometimes = 2; Usually = 3; Always = 4; My health insurance plan was not asked to approve or pay for any kind of care in the last 6 months = NA)

Q42 In the last 6 months, how often were calls to the health insurance plan's customer service taken care of without a long wait? (Never = 1; Sometimes = 2; Usually = 3; Always = 4; I didn't call my health insurance plan's customer service in the last 6 months = NA)

Q43 In the last 6 months, how often did you get all the information or other help you needed when you called the health insurance plan's customer service? (Never = 1; Sometimes = 2; Usually = 3; Always = 4; I didn't call my health insurance plan's customer service in the last 6 months = NA)

Q44 In the last 6 months, how often were the people at the health insurance plan's customer service as helpful as you thought they should be? (Never = 1; Sometimes = 2; Usually = 3; Always = 4; I didn't call my health insurance plan's customer service in the last 6 months = NA)

Office Staff Helpfulness

Q26 In the last 6 months, how often did office staff at a doctor's office or clinic treat you with courtesy and respect? (Never = 1; Sometimes = 2; Usually = 3; Always = 4; I haven't talked to staff at a doctor's office or clinic in the last 6 months = NA)

Q27 In the last 6 months, how often were office staff at a doctor's office or clinic as helpful as you thought they should be? (Never = 1; Sometimes = 2; Usually = 3; Always = 4; I haven't talked to staff at a doctor's office or clinic in the last 6 months = NA)

Global Ratings

Q9 How would you rate your personal doctor or nurse now? Use any number on a scale from 0 to 10 where 0 is the worst personal doctor or nurse possible, and 10 is the best doctor or nurse possible. (Worst personal doctor or nurse possible = 0; Best personal nurse or doctor possible = 10; I don't have a personal doctor or nurse = NA)

Q14 How would you rate the specialist you saw most often? Use any number on a scale from 0 to 10 where 0 is the worst specialist possible, and 10 is the best specialist possible. (Worst specialist possible = 0; Best specialist possible = 10; I didn't need to see a specialist in the last 6 months = NA)

Q37 How would you rate all of your health care from all doctors and other health professionals? Use any number on a scale from 0 to 10 where 0 is the worst health care possible, and 10 is the best health care possible. (Worst health care possible = 0; Best health care possible = 10; I had no visits in the last 6 months = NA)

Q45 How would you rate your health insurance plan now? Use any number on a scale from 0 to 10 where 0 is the worst health insurance plan possible, and 10 is the best health insurance plan possible. (Worst health insurance plan possible = 0; Best health insurance plan possible = 10)

Table 1

Sample Demographic Characteristics

Characteristic	Medicaid		Commercial insurance	
	Latino	White	Latino	White
Gender (% male)	11	9	42	41
Age (% in year-range)				
18–24	28	22	8	4
25–34	39	40	27	18
35–44	24	27	33	29
45–54	6	7	21	32
55–64	2	3	10	15
65–74	1	1	1	2

Table 2

CAHPS 1.0 Means and Standard Deviations for Each Subgroup

Item and dimension	L-Med		W-Med		L-Med		W-Med	
	M	SD	M	SD	M	SD	M	SD
Access to Care								
Q5	1.20	0.46	1.21	0.40	1.18	0.38	1.17	0.38
Q8	2.71	1.22	2.78	1.18	1.84	1.06	1.77	1.06
Q11	2.69	1.14	2.76	1.17	2.86	1.15	2.91	0.98
Q13	1.23	0.43	1.28	0.44	1.26	0.44	1.26	0.44
Q36	3.30	1.03	3.45	0.84	3.32	0.96	3.37	0.93
Timeliness of Care								
Q16	3.34	0.88	3.32	0.87	3.34	0.91	3.41	0.84
Q17	2.91	1.01	2.80	1.01	2.89	1.06	2.91	1.02
Q19	3.16	0.94	3.16	0.96	3.21	0.98	3.20	0.96
Q21	3.12	0.96	3.14	0.97	3.15	1.01	3.15	0.98
Q25	3.01	1.02	3.10	0.98	3.03	1.09	3.06	1.05
Provider Communication								
Q28	3.51	0.74	3.38	0.81	3.45	0.77	3.41	0.75
Q29	3.43	0.84	3.48	0.77	3.51	0.72	3.51	0.69
Q30	3.50	0.75	3.41	0.82	3.50	0.74	3.46	0.73
Q31	3.21	0.89	3.23	0.88	3.28	0.82	3.26	0.83
Q32	3.23	0.92	3.24	0.91	3.28	0.86	3.24	0.83
Q34	3.47	0.84	3.45	0.84	3.50	0.77	3.47	0.78
Health Plan Consumer Service								
Q38	3.51	0.80	3.70	0.66	3.59	0.85	3.57	0.92
Q40	2.89	1.17	2.91	1.15	2.84	1.18	3.02	1.10
Q42	2.82	1.11	2.81	1.09	2.81	1.10	2.83	1.06
Q43	3.19	0.98	3.12	1.01	3.13	0.98	3.11	0.99
Q44	3.28	0.96	3.19	0.96	3.20	0.94	3.16	0.95
Office Staff Helpfulness								
Q26	3.58	0.71	3.59	0.69	3.64	0.66	3.64	0.62
Q27	3.39	0.81	3.35	0.81	3.39	0.78	3.36	0.75
Global Ratings								
Q9	7.98	2.58	7.76	2.58	8.06	1.97	8.02	1.83
Q14	7.94	2.68	7.58	2.82	7.95	2.37	7.96	2.26
Q37	5.38	4.20	7.68	2.42	7.94	3.44	7.83	1.91
Q45	7.84	2.59	7.45	2.00	7.65	2.12	7.31	2.11

Note. Item content can be found in the Appendix. CAHPS = Consumer Assessment of Health Plans Study; L = Latino; W = White (non-Latino); Med = Medicaid; Com = Commercially-insured.

Table 3
CAHPS 1.0 Skewnesses and Kurtoses for Each Subgroup

Item and dimension	Skewness				Kurtosis			
	L-Med	W-Med	L-Com	W-Com	L-Med	W-Med	L-Com	W-Com
Access to Care								
Q5	1.42	1.52	1.67	1.72	0.32	0.02	0.79	0.99
Q8	-0.45	-0.32	0.99	1.11	-1.49	-1.30	-0.37	-0.17
Q11	-0.22	-0.08	-0.34	-0.44	-1.47	-1.49	-1.42	-1.33
Q13	1.00	1.26	1.11	1.09	-0.41	-0.99	-0.77	-0.81
Q36	-1.05	-1.16	-1.20	-1.30	-0.08	-0.28	0.22	0.51
Timeliness of Care								
Q16	-0.98	-0.97	-1.05	-1.19	-0.34	-0.20	-0.17	0.30
Q17	-0.33	-0.44	-0.44	-0.51	-0.97	-1.03	-1.08	-0.89
Q19	-0.80	-0.77	-0.94	-0.95	-0.52	-0.50	-0.34	-0.20
Q21	-0.83	-0.73	-0.88	-0.89	-0.62	-0.41	-0.46	-0.31
Q25	-0.88	-0.76	-0.81	-0.81	-0.55	-0.26	-0.64	-0.62
Provider Communication								
Q28	-1.11	-1.33	-1.18	-1.03	0.81	0.34	0.46	0.24
Q29	-1.39	-1.37	-1.39	-1.29	0.94	1.15	1.32	1.17
Q30	-1.21	-1.30	-1.34	-1.15	0.70	0.55	0.98	0.61
Q31	-0.90	-0.82	-0.86	-0.90	-0.31	-0.13	-0.14	0.10
Q32	-0.95	-0.95	-0.94	-0.85	-0.11	-0.12	-0.06	-0.05
Q34	-1.37	-1.54	-1.43	-1.38	1.35	0.91	1.19	1.08
Health Plan Consumer Service								
Q38	-2.45	-1.64	-2.13	-2.06	5.95	1.91	-3.44	2.82
Q40	-0.55	-0.51	-0.42	-0.71	-1.18	-1.27	-1.37	-0.90
Q42	-0.37	-0.40	-0.38	-0.42	-1.19	-1.21	-1.21	-1.07
Q43	-0.77	-0.83	-0.76	-0.74	-0.68	-0.60	-0.62	-0.66
Q44	-0.86	-1.04	-0.81	-0.80	-0.43	-0.21	-0.56	-0.49
Office Staff Helpfulness								
Q26	-1.66	-1.62	-1.83	-1.72	2.20	1.86	2.84	2.52
Q27	-1.02	-1.09	-1.06	-0.95	0.18	0.17	0.27	0.20
Global Ratings								
Q9	-1.42	-1.16	-1.19	-1.16	1.26	1.67	1.36	1.67
Q14	-1.49	-1.52	-1.45	-1.52	1.41	2.14	1.72	2.14
Q37	-0.25	-1.10	-1.14	-1.10	-1.64	1.22	1.22	1.22
Q45	-1.32	-0.96	-1.09	-0.96	1.12	0.83	1.27	0.83

Note. Item content can be found in the Appendix. CAHPS = Consumer Assessment of Health Plans Study; L = Latino; W = White (non-Latino); Med = Medicaid; Com = commercially insured.

Table 4
Five-Factor Model of CAHPS Report Items: Factor Loadings

Item and dimension	Sample			
	L-Med	W-Med	L-Com	W-Com
Access to Care				
Q5	.43	.51	.51	.46
Q8	-.12	-.15	.24	.17
Q11	.36	.39	.32	.32
Q13	.49	.52	.43	.45
Q36	.61	.66	.55	.58
Timeliness of Care				
Q16	.51	.58	.54	.54
Q17	.42	.52	.42	.51
Q19	.54	.58	.49	.52
Q21	.57	.56	.48	.51
Q25	-.41	-.46	-.34	-.27
Provider Communication				
Q28	.75	.81	.68	.73
Q29	.65	.63	.60	.66
Q30	.75	.81	.68	.72
Q31	.73	.77	.65	.69
Q32	.64	.66	.61	.65
Q34	.50	.53	.43	.53
Health Plan Consumer Service				
Q38	-.08(<i>ns</i>)	-.20	-.15	-.13
Q40	.28	.29	.22	.35
Q42	.46	.52	.51	.54
Q43	.70	.67	.69	.68
Q44	.66	.68	.68	.71
Office Staff Helpfulness				
Q26	.71	.72	.61	.65
Q27	.79	.81	.76	.76

Note. Item content can be found in the Appendix. All items significant at $p < .01$, except as noted. CAHPS = Consumer Assessment of Health Plan Study; L = Latino; W = White (non-Latino); Med = Medicaid; Com = commercially insured.

Table 5
Five-Factor Model of CAHPS Report Items: Correlations Among Factors

Factor and sample	1	2	3	4	5
1. Access to Care					
L-Med	—	.72	.77	.64	.62
W-Med	—	.84	.83	.69	.70
L-Com	—	.85	.89	.71	.69
W-Com	—	.78	.80	.63	.63
2. Timeliness of Care					
L-Med		—	.80	.62	.82
W-Med		—	.78	.63	.77
L-Com		—	.80	.62	.70
W-Com		—	.73	.57	.71
3. Provider Communication					
L-Med			—	.60	.83
W-Med			—	.52	.81
L-Com			—	.56	.79
W-Com			—	.52	.76
4. Health Plan Consumer Service					
L-Med				—	.48
W-Med				—	.47
L-Com				—	.54
W-Com				—	.45
5. Office Staff Helpfulness					
L-Med					—
W-Med					—
L-Com					—
W-Com					—

Note. Item content can be found in the Appendix. All correlations significant at $p < .001$. CAHPS = Consumer Assessment of Health Plans Study; L = Latino; W = White (non-Latino); Med = Medicaid; Com = commercially insured.

Table 6

Multigroup Model Fit Indexes

Model number (and description)	Chi-square	df	NFI	NNFI	CFI	IFI	RMSEA	CI
1. Configurally invariant model of report items (No cross-group equality constraints)	3,137.89	440	.930	.930	.939	.939	.026	(.025, .026)
2. Weak factorial invariance (Cross-group equality constraints imposed on analogous factor loadings)	3,177.96	463	.924	.933	.939	.939	.026	(.025, .026)
3. Partial weak factorial invariance (Cross-group constraints released on 5 pairs of loadings)	3,149.19	458	.930	.933	.940	.940	.026	(.025, .026)
4. Factor correlation invariance (Cross-group constraints imposed on 10 analogous factor correlations)	3,171.72	468	.930	.934	.939	.939	.025	(.024, .026)
5. Partial factor correlation invariance (Cross-group constraints released on 3 of 10 factor correlations)	3,159.93	465	.930	.934	.939	.940	.025	(.025, .026)
6. Higher order model of report items (All 5 reporting factors allowed to load on higher order factor)	3,413.09	468	.924	.928	.934	.934	.026	(.025, .026)
7. Higher order model with global ratings factor incorporated	6,852.22	657	.896	.898	.905	.905	.032	(.032, .033)
Medicaid Samples								
1. Configurally invariant model of report items (No cross-group equality constraints)	1,279.60	440	.967	.975	.978	.978	.018	(.017, .019)
2. Weak factorial invariance (Cross-group equality constraints imposed on analogous factor loadings)	1,319.79	463	.966	.975	.978	.978	.017	(.016, .019)
3. Partial weak factorial invariance (Cross-group constraints released on 3 pairs of factor loadings)	1,296.24	460	.966	.976	.978	.978	.017	(.016, .018)
4. Factor correlation invariance (Cross-group constraints imposed on 10 factor correlations)	1,319.32	470	.966	.976	.978	.978	.017	(.016, .018)
5. Partial factor correlation invariance (Cross-group constraints released on 2 of 10 factor correlations)	1,304.61	468	.966	.976	.978	.978	.017	(.016, .018)
6. Higher order model of report items (All 5 reporting factors allowed to load on higher order factor)	1,650.44	470	.957	.967	.969	.969	.020	(.019, .021)
7. Higher order model with global ratings factor incorporated	3,800.31	658	.926	.934	.938	.938	.028	(.027, .029)

Note. NFI = normed fit index; NNFI = non-normed fit index; CFI = comparative fit index; IFI = incremental fit index; RMSEA = root mean squared error of approximation; CI = confidence interval for RMSEA.