

Measurement of Women's Satisfaction with Primary Care

AHRQ Grant No. R01 HS10237

FINAL REPORT

Carol S. Weisman, PhD¹, Roger T. Anderson, PhD², Sarah Hudson Scholle, DrPH³,
Fabian Camacho, MS², and Jillian T. Henderson, MPH¹

June 2, 2003

¹ Principal Investigator, University of Michigan School of Public Health, Ann Arbor, Michigan (now at Penn State College of Medicine, Hershey, Pennsylvania)

² Wake Forest University, Winston-Salem, North Carolina

³ University of Pittsburgh, Pittsburgh, Pennsylvania

Acknowledgements: The DHHS Office on Women's Health funded the preliminary work for this project through contracts to Drs. Weisman, Anderson, and Scholle (Contract no. 00T00215901D). We would like to thank the staff and patients at the participating clinical sites, and we acknowledge the following individuals involved in data collection and management: Anne Safrit, April Zeoli, Joanna Chapin, Kristine Searcy, Judith Navratil, and Shirley Conti. We also acknowledge Douglas Levine, PhD, for his help with the design of statistical analyses and his contribution toward the development of the final version of the survey tool.

Measurement of Women's Satisfaction with Primary Care

ABSTRACT

This multi-site collaborative project developed and validated the Primary Care Satisfaction Survey for Women, the first survey tool designed to measure adult women's satisfaction with their primary care experiences. Items in the survey were developed using focus groups of women in multiple communities, cognitive testing in several samples of women, and a field test on a multi-site clinical sample of women (n = 1,202). Factor analysis and item response theory (IRT) methods were used to identify three scales: Communication, Administration and Office Procedures, and Care Coordination and Comprehensiveness. The first two scales relate to a specific visit; the third scale relates to health care received in the past 12 months. The scales demonstrate strong psychometric properties (reliability and validity), have structures that are invariant across subgroups of women defined by age and race/ethnicity, and offer greater explanatory power compared to a generic satisfaction measure. The scales also relate to expectations for health care and to overall ratings of health care quality. The 24-item PCSSW may be self-administered or conducted by telephone and may be used in studies to evaluate or improve the quality of primary care for women.

Measurement of Women's Satisfaction with Primary Care

PURPOSE OF THE STUDY

This project developed the first survey tool designed to measure adult women's satisfaction with their primary care experiences. With a growing emphasis on women's health research and on improving the delivery of health services to women, there is a need for patient satisfaction measures that are relevant to women's primary care. This new survey tool should be useful to health plans and other health care organizations interested in evaluating and improving the quality of care for women. It also should be useful for comparing the quality of care for women across subgroups of the population and in different types of organizational settings, including the growing number of women's health programs and centers. (The tool already has been used in studies of the National Centers of Excellence in Women's Health and the Veterans Administration specialized women's health centers.) In addition, because this new tool is based on research on women, it should provide information on patient satisfaction that is relevant and meaningful to women consumers.

BACKGROUND

Rationale

Today, most health care programs use patient satisfaction surveys to assess the quality of care from patients' perspectives. Many surveys are available, some focusing on satisfaction with a specific health care visit or inpatient episode, and some focusing on health care received over a period of time. A key example of the latter is the Consumer Assessment of Health Plans Study (CAHPS) surveys, which are used by the National Committee for Quality Assurance in the assessment of quality of care received in managed care plans during the past year (Hays et al. 1999). Available patient satisfaction surveys, however, were designed for use in general patient populations and were not intended to focus on primary care issues for women (Weisman et al. 2001).

Women make more health care visits than men and are a majority of patients seen in many primary care settings. For that reason alone, information about the quality of their primary care experiences is important. However, women's primary care utilization patterns are complex due to the structural fragmentation of reproductive and general health care (Clancy and Massion 1992). To obtain comprehensive care, many women use two physicians (a generalist and an obstetrician gynecologist) for their regular care (Henderson et al. 2002). Moreover, there are numerous clinical guidelines for the provision of routine preventive services for women (e.g., Pap smears, mammograms, osteoporosis screening) that typically require additional visits. Research also has identified some patient-provider communication issues specific to women's primary care (e.g., better communication on sensitive topics when the physician is female) (Henderson and Weisman 2001). Because generic patient satisfaction tools are designed to be

applicable to all patients regardless of gender, they may not be sensitive to these issues.

As one example, the CAHPS adult survey contains questions asking respondents if they “have *one* person you think of as your personal doctor or nurse [emphasis added]” and if they have seen “specialists” in the last 12 months; these questions do not provide an opportunity for women to report *two* regular physicians (a generalist and an obstetrician-gynecologist), and the question about specialists does not include obstetrician-gynecologists in its definition of specialist. As a consequence, those women who rely on two physicians for their regular health care, or who view their obstetrician-gynecologist as their “personal” physician rather than a “specialist,” may find these questions to be confusing or ambiguous, with unknown effects on their responses.

Existing patient satisfaction tools may be inadequate for measuring women’s satisfaction with alternative models of care delivery or with quality improvement initiatives because generic instruments do not tap into the structure of women’s primary care and neglect some aspects of the process of women’s health care. A tool that captures these dimensions is needed for studies and quality improvement activities in women’s primary care. It is important to recognize that our argument for a patient satisfaction survey tool for women’s primary care is not based on assumptions about women being more or less satisfied with their health care than men. In fact, most research finds no significant mean differences between women and men on generic patient satisfaction tools (Hall and Dornan 1990), although some studies find different *predictors* of satisfaction for women and men (Kolodinsky 1997; Weisman et al. 2000; Weisman et al. 2001). Furthermore, we are not attempting to develop a tool that would enable us to discern gender *differences* in primary care satisfaction. Rather, we are interested in developing a better measure of *women’s* primary care satisfaction. The process of validating a new instrument is lengthy and includes testing the applicability of content found in generic surveys of patient satisfaction as well as developing and testing new items to address areas not covered in existing surveys. The intended result is a valid and reliable patient satisfaction survey that is sensitive to the health care experiences of women.

An alternative approach would have been to develop a set of targeted items unique to women (i.e., not applicable to men) that could be appended to a generic patient satisfaction survey. This approach is taken in the CAHPS surveys, for example, where specific items have been developed for special populations such as Medicaid recipients, children, and persons with specific health conditions. While this approach would preserve the generic item set of the original tool, the limitations would be substantial. It would direct attention only to issues that are unique to women (e.g., reproductive health services) and would not permit reframing existing satisfaction items to be more sensitive to women’s overall experiences. Consequently, the purpose of this project is to develop and validate a tool that reflects the *totality* of women’s primary care experiences, including components of care that are gender-specific and components that women share with men. This tool could then be used as a stand-alone patient satisfaction tool in studies of women’s primary care. That is, the tool is not intended as a substitute to routine use of generic satisfaction instruments to assess average levels of patient satisfaction among both women and men, but rather as a stand-alone tool to collect focused data on women’s health care.

Preliminary work

The idea for this project arose in the Evaluation Working Group of the National Centers of Excellence in Women's Health (CoE). The CoE program is funded by the Department of Health and Human Services (DHHS), Office on Women's Health. All CoEs are located in academic health centers and are required to have clinical care centers that are multidisciplinary and integrate the various components of women's primary care (reproductive and non-reproductive) in women-friendly settings. Charged, in part, with evaluating the quality of care provided in these centers, the Evaluation Working Group sought an appropriate measure of patient satisfaction that could be used in the CoE evaluation project.

The Evaluation Working Group reviewed existing outpatient surveys that include any questions asking respondents to *rate* their health care experiences, as opposed to *reporting* what occurred during health care encounters. (*Ratings* solicit patient's assessments of what occurred, using scales such as excellent-to-poor, very satisfied-to-not satisfied, or numerical rating scales, such as 0-10.) Tools reviewed included the Consumer Assessment of Health Plans Study (CAHPS) commercial and Medicaid surveys; the Picker Institute's Adult Office Visit Survey; the GHAA/Medical Outcomes Study visit rating questionnaire and primary care assessment scales (Davies and Ware 1991; Rubin et al. 1993; Safran et al. 1994); and the Primary Care Assessment Survey (Safran et al. 1998). On the basis of this review, the Working Group concluded that no tool was available that had been designed in whole or in part with a focus on women's primary care. Further, patient satisfaction surveys intended for use by women and men were not necessarily sensitive to the practices (e.g., co-locating routine gynecological and general medical services) that would be expected to improve women's satisfaction with care in the CoEs, compared with standard care.

With initial funding from the DHHS Office on Women's Health, 6 CoEs (University of Michigan, University of Pittsburgh, Wake Forest University, University of California at San Francisco, University of Pennsylvania, and Boston University) conducted a focus group project to identify women's expectations and preferences for their primary care. The focus groups involved 137 women of all ages and diverse racial/ethnic and socioeconomic backgrounds. Diversity of health care experiences was ensured by recruiting participants from local communities rather than through clinical sites. Analysis of the focus group results provided evidence both for new types of survey questions focusing on women's specific health care needs and for reframing standard survey questions to be more sensitive to women's health concerns (Anderson et al. 2001; Scholle et al. 2000).

This preliminary work produced a draft set of survey items for a patient satisfaction survey addressing six aspects of women's health care: accessing care, privacy and comfort, communication with providers, comprehensiveness of care, follow-up care, and overall satisfaction with care. This tool was subjected to additional cognitive testing with 63 women at the University of Michigan and at Wake Forest, as well as 284 survey pretests at Wake Forest. Based on this work, the instrument was revised for clarity of wording and to eliminate items with

very low correlations with global ratings or a large number of “does not apply” responses. In addition, because both the cognitive testing and pretesting conducted for the CoE evaluation project demonstrated that some items tapped women’s experiences with care over time (and across providers) and not just in a specific visit, we divided the items into two categories: items pertaining to a specific visit (16 items) and items pertaining to the overall experience of health care at the site during the past year (17 items). Thus, the test version of the PCSSW included 33 items, and all items were rated on a five-point scale ranging from “not at all satisfied” to “extremely satisfied.”

Three CoEs (Michigan, Pittsburgh, and Wake Forest) formed a consortium to apply for funding from the Agency for Healthcare Research and Quality (AHRQ) to conduct a multi-site field test of the Primary Care Satisfaction Survey for Women (PCSSW) and to assess its psychometric properties (reliability and validity). Although the survey questions had been developed in the preliminary project, a survey tool cannot be recommended for general use until research has determined that it is reliable and valid -- that is, whether the tool measures what it is intended to measure and can be used reliably in the intended population.

The PCSSW tested in this project differs from existing patient satisfaction tools in several ways:

- Some of the items are specific to women (e.g., the ability to get both gynecological and general health care at the same site; the health professional’s knowledge of women’s health issues)
- Other items are new topics not typically included in patient satisfaction surveys but potentially applicable to all patients (e.g., how well office staff keeps the patient informed about waiting time; the health professional’s interest in the patient’s mental and emotional health).
- Other items are similar to items in generic patient satisfaction surveys but are worded to be consistent with women’s framings as discovered in the focus groups (e.g., the health professional’s ability to answer questions in a “sensitive and caring way”).

In addition, all items in the PCSSW are worded to refer to the patient’s “health professional(s)” rather than to a “regular doctor” or to “your doctor,” due to previous research showing that many women rely on more than one “regular” physician for their health care and to take into account the role of nurse practitioners and physician assistants in primary care.

RESEARCH QUESTIONS

The specific aims of this project were to complete the development of the PCSSW and to:

- (1) Assess the reliability and validity of the PCSSW in a large field test of women and demonstrate the performance of the survey tool relative to standard patient

satisfaction surveys;

- (2) Assess the comparability of patient satisfaction in subgroups of women defined by age and race/ethnicity; and
- (3) Analyze the variables associated with satisfaction with primary care in a large, geographically and sociodemographically diverse sample of women.

An additional research question was addressed because the national evaluation of the CoE program was being conducted simultaneously with this project. In the CoE evaluation project, the PCSSW was used to test the hypothesis that women served in women-specific primary care programs (the CoE clinical care centers) report higher satisfaction with care than a community sample of women. The methods and results for this research question are not reported here, but are available in Anderson et al. (2002).

OVERALL STUDY DESIGN AND DATA COLLECTION

Study design

The study consisted of a survey of 1,202 women making primary care visits at sites affiliated with the three participating health centers (University of Michigan Health System in Ann Arbor, Michigan; University of Pittsburgh Medical Center in Pittsburgh, Pennsylvania; and Wake Forest University Baptist Medical Center in Winston-Salem, North Carolina). The use of multiple sites in different parts of the country ensured a diverse sample of women seeking health care in different types of practice settings. The women completed self-administered questionnaires in the sites before and immediately after a primary care visit. We defined a “primary care visit” as a visit for a routine checkup, gynecological exam, prenatal care, acute care, or routine follow-up care with a doctor or other independent practitioner (advanced practice nurse or physician assistant). Excluded were emergency visits and expedited visits to drop off a lab specimen or to receive a single procedure such as a flu shot, allergy shot, or contraception injection.

The pre-visit questionnaire included questions on subjects’ patterns of health care utilization, expectations for the visit, health status, and sociodemographics. The post-visit questionnaire included questions on what occurred at the visit (e.g., services received, providers seen), satisfaction with the visit (including the PCSSW and comparison patient satisfaction items), behavioral intentions, self-efficacy for preventive care, and additional sociodemographics.

Sites

Subjects were recruited at primary care sites affiliated with the three participating health centers. Located in different geographic regions, these centers provide access to diverse primary care settings and patient populations. In all sites, a mix of personnel – including residents,

primary care and specialist physicians, and advanced practice nurses or physician assistants – provided primary care services.

At the University of Michigan, subjects were recruited from three clinical areas (internal medicine, family practice, and obstetrics and gynecology) in one ambulatory care facility in Ann Arbor. The facility is affiliated with the CoE at the University of Michigan Health System and is staffed by physicians, nurses, and certified nurse midwives. Patients served reflect the sociodemographic composition of the community and are predominantly white and privately insured. With the exception of obstetrics-gynecology, these clinical areas serve both women and men; survey days and hours were selected to capture high concentrations of female patients.

At the University of Pittsburgh, sites included a primary care clinic that was a major practice site for faculty and resident physicians in the Division of General Internal Medicine, and a large obstetrics and gynecology clinic at Magee-Womens Hospital affiliated with the CoE program. The latter is staffed by resident physicians as well as nurse practitioners and certified nurse midwives. These primary care sites serve a population diverse in race/ethnicity and income, and they offer several half-day sessions devoted to a women's health program. Survey days were selected to capture a sample with a high representation of patients of lower socioeconomic status and African American ethnicity.

At Wake Forest University Baptist Medical Center, subjects were recruited in Winston-Salem from an academic multi-specialty practice for faculty and resident physicians affiliated with the CoE program and from two freestanding family practice satellite clinics. For the latter, one clinical site served a mostly African American residential area of the city, and the other served an affluent suburban area that was predominantly Caucasian. The academic clinic holds daily primary care clinics and offers several sessions per week focused on women's health topics. Providers at each of the family health care clinics included a team of approximately four full-time physicians, a physician's assistant, and a family practice nurse. Access to obstetrics and gynecology as well as other specialty care occurs by referral to plan-affiliated practices.

Data collection procedures and sample

Women were eligible to participate in the study if they were ages 18 and over, English-speaking, not known or suspected to be cognitively impaired, able to complete the questionnaire without assistance or proxy, and making a primary care visit at the time of the survey.

Trained project personnel recruited the participants and obtained written informed consent. Subjects were approached when checking in for the visit and invited to participate in the study on a "next available" basis. In this approach, the recruiter sought the next eligible subject who had completed checking in for the visit and could be approached within 2-3 minutes of being seated in the waiting room and before being called in for the visit. Eligible volunteers completed a consent process that informed them that their responses were both confidential and anonymous as their name or other personal identifying information would not be collected. Participants completed both a pre-visit and post-visit questionnaire at the site. Upon completion

of both questionnaires, subjects were compensated for their time in cash (\$20) or equivalent-value coupons from local vendors.

Data from all sites were pooled for analysis. The final sample of 1,202 (400 or 401 at each site) completed surveys represents an overall participation rate of 69% of eligible subjects. Participation rates varied by site: 56% at Wake Forest, 72% at Michigan, and 88% at Pittsburgh. The reasons for declining participation were not having time to remain after the visit to complete the post-visit questionnaire (55% of non-respondents), lack of interest (24%), too ill to participate (14%), and no reason given (7%).

Table 1 shows characteristics of the study sample. The average age of respondents was about 42 years. About two-thirds of respondents were white, non-Hispanic, and 23.5% were black, non-Hispanic. Education and household income varied widely, with one-third of women having a high school education or less, and 19.5% having post-college education. While 26% had household incomes below \$20,000, 17% had incomes above \$75,000. With regard to health care utilization, respondents made an average of 7.8 visits per year (standard deviation = 8.2), and 76% considered the site where they were sampled to be their usual source of care. Patients visited the sites for a variety of reasons including routine exams (26%), prenatal care (16%), a new health problem (27%), and follow-up care (32%).

SCALE DEVELOPMENT: The PCSSW

Methods

We used a three-step process employing factor analytic methods and item response theory (IRT) to select the final version of the PCSSW from the test set of 33 items (Horn 1965; Hu and Bentler 1995; Kaiser 1970; Samejima 1997; Velicer 1976). First, an exploratory factor analysis was performed on half of the sample selected at random (with the remaining half serving as a test set in the confirmatory analyses described below). Because the items stemmed from different content domains, the analyses were conducted separately for items referring to a specific visit and for items referring to care over the past year. Four items with excessive missing values or excessive skewness were removed before these analyses were conducted: (1) the health professional's ability to make me feel comfortable during a gynecological (pelvic) exam, (2) the health professional's comfort talking about sensitive issues like sexuality, (3) the health professional's comfort talking about natural or alternative therapies, and (4) child care here if I need it.

The principal factor method with squared multiple correlations as prior communality estimates was used to extract the factors, and an oblique solution was obtained using the promax factor rotation technique. The number of factors to retain was investigated by examining the scree plots, using Horn's parallel analysis criterion (Horn 1965), conducting Velicer's minimum average partial procedure (Velicer 1976), and noting the number of eigenvalues above the average value.

Second, an item elimination and selection process was performed for each factor in order to develop indicative scales. Items were assessed by considering the magnitude of their factor loadings, assessing the item-total correlation, and fitting Samejima's graded IRT model (Samejima 1997) to each set of candidate items. IRT is a method for characterizing the relationship between a person's responses to specific survey items and her standing on an overall construct. IRT models provide Item Characteristic Curves (ICC), also referred to as trace lines. These parametric curves describe the probability (on a scale of 0 to 1.0) that a particular respondent will choose a given item response category given her underlying satisfaction level, where satisfaction is conceptualized as an unbounded continuous latent variable with a mean of zero and a standard deviation of one. By considering the parameters that define these curves, it becomes possible to discover items that discriminate poorly between respondents with differing degrees of satisfaction. Samejima's model introduces constraints to the trace lines and assumes the item categories are on an ordinal scale. The trace line for the lowest category will approach a probability of 1.0 as the respondent's satisfaction level increases in the negative direction, and it will approach zero as the respondent's satisfaction increases positively. Conversely, the trace line for the highest category will approach a probability of 1.0 as the level of satisfaction increases, and will approach zero as satisfaction decreases. Categories in between are constrained to have trace lines that reach a peak and decrease to a probability of zero in either direction.

The point where an item response category trace line would reach 1.0 at a particular level of satisfaction indicates a high level of consistency between levels of a specific item response and satisfaction. The ideal for ordinal response sets, like those in the PCSSW, would be for respondents with low satisfaction to have a high probability of selecting low category responses and for respondents with high satisfaction to have a high probability of selecting high category responses. The set of parameters that determine the trace lines are reported in Table 3. The discrimination index (denoted as "a" in Table 3) is a measure of how well the item response continuum may differentiate levels of the latent construct (satisfaction). Items with low index values have substantial overlap among response category trace lines. Items with high index values have trace lines with little or no overlap.

Other parameters derived from the trace lines are called location parameters (b), and the number of location parameters is equal to the number of response categories minus 1. These location parameters assess item difficulty, which is defined as the point along the measurement of the latent variable (satisfaction) where respondents are likely to choose a response. By convention, the parameter "b₁" denotes the point along the continuum for which there is a 50% probability of selecting the lowest response; "b₂" denotes the place where there is a 50% probability of selecting either the lowest or next-to-lowest response; and so on (as shown in Table 3). Thus, items with high b₁ parameters are better at discriminating among respondents with low satisfaction.

In addition to the ICC of the individual items, IRT modeling assesses the contribution of each item in determining the precision with which patient satisfaction can be measured. Unlike classical test theory, in which precision is viewed as associated with item reliability, IRT views

precision as a function of the level of the latent construct (satisfaction) itself and can be extracted from the IRT model by considering an information function plotted as an Information Curve. The concept of “information” is akin to the certainty with which the underlying construct is being measured. Satisfaction items with high information, compared with satisfaction items with low information, contribute more to the certainty (precision) with which satisfaction is being measured within the scale. The information of the overall scale is the reciprocal of the standard error of the estimate of theta. At a given satisfaction level, items with higher information contribute more to the overall precision of the scale. Information curves can be used to indicate which items are providing a high level of information and therefore should be retained in the scale. However, items with lower information are not necessarily candidates for removal because they provide at least some information and may perform well on other aspects of psychometric testing, such as item discrimination.

The final selection of items was based on consideration of the quantitative and qualitative characteristics of the items and the subscales. That is, we wanted to include items that were highly reliable, that were sensitive to the full range of the latent variables, and that contributed to the overall reliability of each subscale. At the same time, we wanted items that reflected the full range of conceptual content in each of the domains. The final set of items was selected based on all of these considerations.

Third, once having settled on the items for each scale, a new factor analysis was repeated on the exploratory sample to obtain the factor loadings of the new model. In order to assess the stability and generalizability of the proposed factor structure, the same analysis was conducted on the test sample, and results were compared. The χ^2 test and the Tucker Reliability Index (Hu and Bentler 1995) were also assessed to evaluate the fit of the new factor model on the test set.

Results

The results of the analysis of visit-specific items will be discussed first, followed by the results on items rating experiences during the past year. **Table 2** shows the results of the exploratory factor analysis. In the analysis of the visit-specific items, we decided to retain two factors based on procedures to determine the number of factors. The two oblique factor solutions explained approximately 70% of the variance in the dataset.

Ten items loaded on the first factor, Communication. **Table 3** shows the estimated parameters from the IRT models. The discrimination parameters suggest that all items are indicative of their corresponding scale, although some items possess better discriminating ability than others. Items Q11i, Q11j, Q11k, Q11l, and Q11h form a cluster with high discriminating properties. Item Q11i (“the health professional’s ability to explain things clearly”), in particular, has a high slope parameter ($a=5.09$), which indicates that it discriminates well among respondents with high versus low satisfaction. It also has the largest negative location parameter ($b_1 = -2.42$), which suggests that this item is the best at discriminating among respondents with low satisfaction. As an example, **Figure 1** shows the ICC curves for item Q11i. **Figure 2** shows

the information curves for the Communication Scale items, with a cluster of similar performing items distinguishable from less reliable items. This model strongly suggests that these items should be retained as components of a scale assessing a trait identified as satisfaction with communication.

Of the ten items that loaded on the Communication factor, eight items were retained in the final scale. Based on the pattern of item-total correlations, factor loadings, and ICCs, as well as the face validity of the items, two items were dropped: Q11n (“The chance to get everything I need at this visit”) and Q11m (“My health professional’s knowledge of my medical history”). Generally, these items had lower loadings on the initial factor, were highly correlated with other items, and performed less well in the IRT analyses; they also did not fit as well conceptually with other items in the scale. Several items that performed less well in the analyses were included in the final scale because of their importance to the content validity of the scale.

A second scale, Administration and Office Procedures, was formed with six items loading on the second factor identified among the visit-specific items (see Table 2). One item that loaded equally on both of the two initial factors, Q11f (“The chance to talk to my health professional with my clothes on”), was grouped with the second factor. This item was retained because of its conceptual importance for this tool (as evidenced by the focus group results) and placed with the Administrative and Office Procedures scale because it had marginally better performance and was more interpretable with the items on visit procedures. **Table 3** shows that the items in this scale have somewhat lower discriminating ability compared to the Communication items. Item Q11f, in particular, does not possess as much discriminating ability as the others (-1.79), although it also has the largest negative location parameter ($b_1 = -2.60$) and thus appears to be best at distinguishing among individuals with low satisfaction. **Figure 3** shows the information curves for the Administration and Office Procedures items and illustrates this pattern.

For the items pertaining to health care during the past year, the factor analysis suggested a single factor solution. This factor explained about 64% of the variance in the item set. The initial factor analysis identified three items with somewhat lower loadings (Q12j, Q12k, and Q12m), and the IRT model suggested that three items (Q12j, Q12k, and Q12l) performed less well in terms of their discriminating ability. (Refer to **Table 3** and **Figure 4**.) We decided to drop Q12j (“the chance to see the same health professional at each visit”) and Q12k (“the chance to see a health professional of the gender I prefer”) based on the inferior psychometric properties. Another item, Q12g (“how well my health information is kept private”) was dropped because it was less relevant to the construct despite its reasonable psychometric properties. Several items that we considered to be critical for the content validity of the scale based on our focus groups were retained despite their marginal performance in the psychometric tests. For example, Q12l, “the chance to get both gynecological and general health care here,” had a high loading (.81) in the exploratory factor analysis. The IRT analyses suggested that this item had lower overall discriminating ability but may be better at discriminating among those with low satisfaction. This item was retained in the scale because of its importance for content validity. Finally, ten items were selected for the Care Coordination and Comprehensiveness Scale.

After deciding on the items for each scale, a confirmatory factor analysis was performed on both the random half of the sample used in the exploratory factor analysis and the remaining test sample (**Table 4**). This showed that the factors explained 70% of the total variance in the visit-specific items and 66% of the variance for the past-year items. The Tucker and Lewis reliability coefficient was 0.94 for the visit-specific item set and 0.86 for the past-year item set, suggesting excellent to good model fit. Comparison of the factor loadings from the initial sample to the test sample provided evidence that the factors were stable, although the second factor loadings for the visit-specific item set exhibited a relative alteration, suggesting that reliabilities may be lower for this scale than the other.

Table 5 presents the final scales with the means and standard deviations for each item. The two scales measuring visit-specific satisfaction include the 8-item Communication Scale and the 6-item Administration and Office Procedures Scale. The 10-item Care Coordination and Comprehensiveness Scale measures satisfaction with health care during the past 12 months. Each PCSSW item is rated on a five-point scale: 1 = “not at all satisfied,” 2 = “somewhat satisfied,” 3 = “satisfied,” 4 = “very satisfied,” and 5 = “extremely satisfied.” For the validity analyses, a score for each scale was calculated by summing the items, with the respondent’s scale mean for non-missing items imputed if there were fewer than 25% missing items on the scale.

The final version of the PCSSW is shown in **Appendix A**.

SPECIFIC AIM 1: METHODS AND RESULTS

Methods

Several approaches were used to assess the reliability and validity of the PCSSW scales. For each PCSSW scale and two comparison generic patient satisfaction measures (the Medical Outcomes Study Visit Satisfaction scale and a CAHPS overall rating item, described below), we present the mean and range. In addition, for the PCSSW and MOS scales, we show the percent with the highest possible rating (as a measure of a potential ceiling effect) and the coefficient alpha as a measure of internal consistency reliability. To assess convergent validity with existing generic patient satisfaction tools, we calculated the correlation of the PCSSW scales with the MOS scale (a visit-specific scale) and the CAHPS item (which refers to care during the past year), using unadjusted Pearson correlations. To assess discriminant validity, we computed means for each of several known groups (adjusted for site, age, and perceived health status) and p-values from t-tests comparing the means. In addition, we present the proportion of variance in the criterion item explained by the satisfaction item, as measured by eta-squared (η^2). Eta-squared is the proportion of the sum of squares attributable to the criterion variable divided by the total corrected sum of squares (Becker 1999). An eta-squared close to 0 implies that the two groups are difficult to distinguish, and an eta-squared close to 1 implies there is a clear difference between the two groups. Predictive validity was assessed in the same manner.

To compare the PCSSW scales' ability to capture the variance in quality assessments to generic scales, we conducted linear regressions with the overall visit quality rating and the CAHPS rating of the overall quality of care during the past year as the dependent variables. In the linear regression models, we first entered site and patient covariates (age, education, and perceived health status) and then the satisfaction scales, with separate regressions for the generic MOS Visit Satisfaction Scale and for the PCSSW scales. The p-value and proportion of the variance explained represents the contribution of each satisfaction scale.

All analyses were conducted using SAS Version 8.1.

Measures for Validity Analyses

Additional measures used in the validity analyses are described below.

Generic satisfaction tools. We used three generic measures of satisfaction with outpatient care for our convergent validity comparisons: the Visit Satisfaction scale from the Medical Outcomes Study (MOS), the rating of the quality of all health care during the past year from the Consumer Assessment of Health Plans Study (CAHPS), and a single item on the quality of care at the visit.

The *MOS Visit Satisfaction* scale is a multi-item visit satisfaction measure. It consists of nine items tapping multiple dimensions of care (e.g., access, time spent with provider, communication, technical quality, interpersonal quality) as well as a rating of the overall visit. It uses a 5-point excellent-to-poor response set (rated 5 to 1) and is summed to give a score. The instrument has been widely used in research on primary care and has been found to discriminate between types of practice settings and health plans and to predict such outcomes as returning for follow-up care and whether patients will change physicians within six months (Davies and Ware 1991; Rubin et al. 1993).

The *CAHPS item* is a single-item rating of the quality of care during the past year. The item asks respondents to rate "all of your health care in the last 12 months from all doctors and other health professionals at this office or clinic." The item is rated on a scale from 0 (worst health care possible) to 10 (best health care possible). This item is used for benchmarking satisfaction in health plans and typically is reported as percent of enrollees who rate their plan with a score of 8 or higher (NCQA 1998). In the context of this project, this item reflects the woman's overall assessment of the quality of care at the site in the past year.

Finally, for comparing the variance in overall satisfaction ratings explained by a generic visit-based scale (the MOS) and the PCSSW scales, we also used a single-item rating of the *overall quality of care at the visit*: respondents rated their satisfaction with the "overall quality of care at this visit," using a five-point response set (1 = "not at all satisfied" to 5 = "extremely satisfied").

Continuity of care. Having a regular source of health care or a regular provider (i.e., site and provider continuity) is known to be associated with higher levels of patient satisfaction (Aharony and Strasser 1993; Cleary and Mc Neil 1988; Donaldson 2001). Measures of these constructs were used to define known groups for assessment of discriminant validity. Women's patterns of continuous care at the site and with a regular health care professional were measured as follows: (1) for those using the site as their usual source of care, having longer tenure at this site (two years or longer), versus shorter tenure (less than two years); and (2) for women who have a regular health care professional, seeing the regular health professional at the visit, versus seeing another health care professional. We hypothesized that women who had longer tenure at their usual source of care would have higher ratings on both the Communication and the Care Coordination and Comprehensiveness scales (as well as the MOS Visit Satisfaction scale and the CAHPS rating). We also hypothesized that women who saw their regular doctor on the day of the visit would have higher ratings on the Communication scale (as well as the MOS scale) compared with women who did not see their regular doctor (women without a regular provider were excluded).

Comprehensiveness of care. The comprehensiveness of preventive care, as an important aspect of the technical quality of care, was hypothesized to be associated with higher satisfaction with care. Although studies linking comprehensiveness of care with patient satisfaction are sparse, the hypothesis is supported by some literature (Cleary and McNeil 1988; Orlando and Meredith 2002; Schauffler et al. 1996; Sitzia and Wood 1997). Three variables reflecting comprehensiveness were defined for the discriminant validity analyses. *Comprehensiveness of preventive screening services* was defined as the number of age-appropriate screening services provided at the visit. For all women, these included blood pressure check, Pap smear, and physical breast exam. For women ages 50 and over, mammogram, blood cholesterol test, and colon cancer screening also were included. The sum of services received was dichotomized for analysis: a high score was three or more services, and a low score was 0-2 services. *Comprehensiveness of preventive counseling* was defined as the number of age-appropriate counseling topics discussed during non-illness visits. For all women, topics were: smoking or quitting smoking; nutrition or diet; physical fitness or exercise; alcohol or drug use; calcium intake and risk of osteoporosis; violence in the home or family or relationship problems; sexual function or problems; work or financial problems; stress management; alternative therapies, such as herbal products or massage therapy. Additional topics for women ages 18-44 included preventing unwanted pregnancy or planning a pregnancy. Additional topics for women ages 45 and older also included menopause or hormone replacement. The sum of counseling topics was dichotomized for analysis: a high score was counseling on at least one topic at the visit, and a low score was no counseling. Finally, we asked women to rate the *completeness of their visit*: "At your visit today, did you get everything that you thought you needed?" This was scored yes or no. We hypothesized that women who had more comprehensive services based on each of these measures would have higher ratings on both visit-based and past-year measures of satisfaction.

Behavioral intentions. Measures of behavioral intent are viewed as determinants of subsequent behavior and are often used in patient satisfaction studies as proxies for post-visit

outcomes when these outcomes cannot be observed directly. Patient satisfaction has been found to be correlated with intention to return to the provider, adhere to providers' recommendations, recommend the provider to others, and similar items (Aharony and Strasser 1993; Ware and Davies 1983; Ware and Hays 1988). To assess predictive validity using behavioral intentions, we asked women to respond "definitely yes," "probably yes," "probably not," and "definitely not" to the following four questions: "Thinking about any health care advice or recommendations you received today from your health professionals do you plan to follow this advice?" "Based on your experience at this visit, do you plan to return to this office or clinic for care?" "Based on your experiences at this visit, would you want to see the same health care professional you saw today again?" "Based on your experiences at this visit, would you recommend this office or clinic to a family member or friend?" We compared satisfaction scores of women who rate each item "definitely yes" versus all other responses. We hypothesized that women with definite intentions to follow the health professional's advice from today's visit and to want the same health care professional from today's visit would have higher visit-based satisfaction scores. Further, we expected that women with definite intentions to return to the office or to recommend the office would have higher scores on both the visit-based and past-year ratings of satisfaction.

Self-efficacy for preventive care. We hypothesized that women who received higher quality preventive care and who were more highly satisfied with their care should have higher confidence in their knowledge of preventive care. The relationship of satisfaction levels to health-related outcomes has not been well studied (Aharony and Strasser 1993; Cleary and McNeil 1988), but a logical hypothesis for primary care is that a higher quality of both the technical and process aspects of care (especially communication between providers and patients) should result in both higher satisfaction levels and higher levels of patient knowledge and self-efficacy for health promotion and disease prevention. Knowledge confidence is theoretically an important goal of patient education. Bandura, Adams, and Beyer (1977) specified that vicarious learning or modeling and verbal persuasion (e.g., encouragement and support from others) are important precursors of self-efficacy and behavioral adherence to recommendations. The extent to which primary care providers enable and support knowledge confidence should therefore reflect the quality of directed discussion of prevention, motivation and importance placed upon the behavior, and patient activation to engage in seeking preventive screening.

To measure knowledge confidence of preventive care, we used six items to measure how sure women were that they knew when to have a mammogram, when to have the next Pap smear, how often to do a breast self-exam, when to have the next gynecologic exam, when to get a cholesterol test, and the steps to take to prevent getting osteoporosis. Each item was scored on a four-point scale from "very sure" (4) to "not at all sure" (1). Internal consistency reliability for the item set is acceptable ($\alpha = 0.82$), so the items were summed. Women who scored 20 or higher on this scale were compared with those with lower scores. We hypothesized that women with greater self-efficacy would have higher ratings on the PCSSW Communication scale and the MOS scale.

Results

Reliability and Convergent Validity of the PCSSW: Each of the PCSSW scales has high internal consistency, with coefficient alpha of 0.96 for the Communication Scale, 0.88 for the Administration and Office Procedures Scale, and 0.95 for the Care Coordination and Comprehensiveness Scale (**Table 6**). The Communication Scale appears to be more subject to a ceiling effect than the other PCSSW scales and the MOS Visit Satisfaction Scale.

The PCSSW has very good convergent validity. Correlations with the MOS Visit Satisfaction Scale (0.67 to 0.73, $p < .001$ for all) and with the overall quality of visit rating (0.60 to 0.74, $p < .001$ for all) are somewhat higher than correlations with the CAHPS overall rating (0.42 to 0.61, $p < .001$) (**Table 7**). The Administration and Office Procedures scale has the lowest correlations with the generic measures.

Discriminant Validity of the PCSSW: We examined the ability of the PCSSW scales to distinguish among groups generally expected to have different levels of satisfaction based on previous literature, and compared this discriminant validity to that of the MOS scale and CAHPS item (**Table 8**). The Care Coordination and Comprehensiveness Scale was able to discriminate among groups based on length of time at the clinical site, with a mean of 38.7 for women who had used the site for two or more years, compared with 37.6 for those with a shorter tenure at the site ($p = .02$), though the proportion of variance in length of time explained by the scale was minimal. All of the measures showed significantly higher satisfaction ratings when the regular doctor was seen at the visit.

All three PCSSW scales and the generic tools showed large differences in satisfaction based on the comprehensiveness of the visit. For example, women who said they “got everything they needed today” at their visit had an adjusted mean score of 33.5 on the PCSSW Communication Scale, compared with 24.3 for women who said “no” to this question ($p < .0001$), with this scale explaining 9% of the variance in whether women agreed or disagreed that they “got everything they needed” at the visit. Women who received counseling on preventive topics also had higher satisfaction ratings on all PCSSW scales and the generic tools, with the Care Coordination and Comprehensiveness Scale having the largest effect size. Both the Communication and the Care Coordination and Comprehensiveness Scale differentiated among women who did and did not receive at least three preventive screening services at the visit.

Predictive Validity of the PCSSW: All of the satisfaction measures were significantly associated with women’s behavioral intentions and self-efficacy for preventive care (**Table 9**). The Care Coordination and Comprehensiveness Scale had the strongest effect on women’s intention to recommend the office/clinic to others; the Communication Scale had the highest effect on women’s intentions to follow advice received at the visit and on their desire to see the same health care professional again.

Contribution of the PCSSW to Overall Quality Ratings: In regression analyses examining the relationship of the generic MOS Visit Satisfaction scale and the PCSSW scales to overall ratings of the quality of care, two of the PCSSW scales (Communication and Care

Coordination and Comprehensiveness) contribute greater explanatory power than the generic MOS Scale (**Table 10**). For example, the Communication Scale explains 75% of the variance in the overall visit quality rating, with patient characteristics entered into the model, compared with 50% of variance explained by the MOS scale. The Care Coordination and Comprehensiveness Scale explains 59% of the variance in overall visit quality and 42% of the variance in the overall CAHPS rating, compared with 50% and 35% of variance explained, respectively, by the MOS scale.

SPECIFIC AIM 2: METHODS AND RESULTS

Methods

We also used multiple group structural equation modeling to compare the similarity of the factor structure across different race/ethnic groups (white, black, and other) and age groups (18-35, 36-54, and 55 and over) that could be defined in this sample. Although the development of the items for the PCSSW was conducted in focus groups stratified by age group and by race/ethnicity to ensure that the items were meaningful to women of all ages and ethnicities, testing for factor invariance provides evidence of PCSSW properties for these subgroups. The form of the structural equation model is shown in **Figure 5**, which graphically illustrates a set of linear relationships between observed (represented by boxes) and latent, hidden, or error variables (represented by circles) suggested by the factor analysis. Straight arrow links represent linear relationships between modeled variables and correspond to regression coefficients, labeled in the figure as beta1 through beta24. Curved arrows represent covariances between variables.

In order to test factor invariance across different subgroups, the models were simultaneously fit to each group with the regression coefficients between the latent scales and the observed variables constrained to be equal across groups. The parameters of the model were estimated by minimizing a maximum likelihood discrepancy function. Fit indices were then evaluated and also compared to the fit indices of the unconstrained model. Good model fit and performance when compared to the unconstrained model would indicate that there is not sufficient evidence to reject the hypothesis of factor invariance (equal regression weights across groups).

Results

For the multiple group model fit on the different race/ethnic subgroups, the minimum discrepancy score divided by the degrees of freedom (CHISQ/DF) was 5.0, slightly above the recommended rules of thumb of 2 or 3 (Carmines and McIver 1981). The Bentler-Bonnet normed fit index (NFI) was .964, and the Comparative Fit Index (CFI) was .971, above the recommended .90 rule of thumb (Bentler and Bonnet (1980), suggesting fit was good enough that it cannot be improved substantially above the baseline independence model. The root mean square error of approximation (RSMEA) was .06, slightly above the recommended value of .05, the suggested threshold indicating a model with a reasonable error of approximation (Browne and Cudeck 1993). The unconstrained multiple-group model shows an almost identical fit

(CHISQ/DF = 5.2, NFI = .965, RSMEA = .060) with a very trivial decrease in Aikaike Information Criteria (.01%) when compared to the constrained model. Taken together, these results support factor invariance for the race/ethnic groups at the current level of precision afforded by the sample, with deterioration in model fit more likely due to slight model misspecification than to the addition of constraints.

Similarly, the multiple group fit on the different age groups had a CHISQ/DF of 5.11, a NFI of 0.964, a CFI of .970, and an RSMEA of .06. The unconstrained model had a similar fit (CHISQ.DF = 5.37, NFI = 0.964, CFI = .971, RSMEA = .06), with a trivial increase in the Aikaike Information Criteria (0.5%). These results also suggest factor invariance across age groups.

Table 11 shows the estimated factor loadings for different groups and may be used to descriptively assess the extent of heterogeneity across race/ethnic and age categories, though from the previous analysis we believe that observed variation could be due to sampling error rather than systematic differences. Descriptively speaking, the greatest variability is observed across race/ethnic groups in the factor loadings of item Q12l (ranging from 56 to 78); this variation suggests that access to both gynecological and general health care accounts for a higher proportion of satisfaction with Care Coordination and Comprehensiveness among African American women.

SPECIFIC AIM 3: METHODS AND RESULTS

Framework

We examined predictors of patient satisfaction using the PCSSW by testing a path analysis model that was constructed based upon the theoretical framework underlying the instrument's development. In this framework, we modeled three types or classes of variables, which we refer to as tiers.

In the first tier are exogenous variables such as age, perceived health status, and measures of dependence on a health care center or practice for health care. The latter includes both continuity of provider (whether the woman has one regular provider, two, or none) and continuity of site (getting one's usual health care from the site of the index visit). This tier also includes our theoretically derived "discrepancy score," which assesses the extent to which one's expectations for the visit are or are not met. (These measures are described in detail below.) These background variables are tested as conditioning variables that have either direct or indirect effects on patient satisfaction.

The second tier consists of the three PCSSW scales, our conceptually distinct components of patient satisfaction, which underlie global ratings (our third tier). We modeled the three PCSSW scales to identify how variables in the first tier predict or correlate with each component. The third tier consists of two global ratings of the quality of care that are intended to summarize all experiences at the visit and over the past year. The CAHPS rating (described

above) is the global measure of quality of care in the past year. The global rating for the visit is based on one item from the MOS Visit Satisfaction scale: a five-point rating (ranging from excellent to poor) for “this visit overall.”

Measures

The additional variables included in these analyses are described below.

Expectations discrepancy: Patients have expectations or goals for their health care visits and a desire to satisfy those goals, and they frequently make requests or verbalizations of those desires (Bell 2001). Expectations may refer to specific aspects of care (e.g., comprehensive services in one location) as well as to the manner in which care is provided (e.g., clear explanations by the provider). Patient’s expectations for care are derived from multiple sources, including past experiences (Kravitz 1996), and may vary by demographic characteristics such as age, ethnicity, and socioeconomic status (O’Malley et al. 2000). Satisfaction with care received often is hypothesized to be a function of met expectations for health care experiences, but this assumption is rarely tested and when tested yields mixed results (Cleary and McNeil 1988; Peck et al. 2001; Sitzia and Wood 1997).

In this project, because we had data both before and after an index primary care visit, we were able to construct a measure of met expectations for the visit. To measure met expectations, we constructed a discrepancy score based on comparing expectations for high quality care before the visit and the actual ratings of care following the visit. Expectations and ratings were measured using five questions tapping some key dimensions of primary care: “talking to the health professional with my clothes on” (privacy); “getting everything I need at one visit” (comprehensiveness of care); “getting help scheduling my next appointment” (coordination of care); “having a health professional who includes me in decisions about my care” (decision making); and “having a health professional who coordinates all the health care I receive” (coordination/continuity). Expectations were measured in terms of how important each of the dimensions was to the patient prior to the visit; after the visit, the patient indicated the extent to which the dimension was experienced during the visit. A score was derived by taking the mean discrepancy for the five items combined. For each item, the discrepancy score could vary from – 2 to +2, where 0 indicates an exact match of expectations and experiences, –2 indicates the largest negative dissonance (where actual experience was two response category units below expectations), and +2 indicates the largest positive dissonance (where actual experience was two response category units above expectations). Therefore, the hypothesized direction between the expectations discrepancy score and satisfaction with care is positive. A caveat of this approach is noteworthy: although this discrepancy score measure is theoretically based, it has not been tested for concept validity and reliability.

Site continuity: Continuity of care at the site was measured as the proportion of all health care visits in the past 12 months that were made to this clinical site. This is similar to continuity measures that focus on specific providers rather than sites of care (Gill and Mainous 1998; Magill and Senf 1987). The proportion ranges from 0.04 to 1.00.

Provider continuity: This is a three-category variable indicating whether women have no regular primary care provider, one regular provider (a generalist physician), or two regular providers (both a generalist physician and an obstetrician-gynecologist). Seeing both a generalist and an obstetrician-gynecologist, compared with a generalist alone, has been shown in previous research to be associated with receiving more comprehensive preventive care (Gallagher et al. 2001; Henderson, Weisman, and Grason 2002; Weisman 1996; Weisman and Henderson 2001). However, seeing two physicians for regular care is, by definition, an indicator of less provider continuity, and it has not been found to be related to patient satisfaction levels, using other measures of satisfaction than those reported here (Weisman and Henderson 2001). Because women who have one regular provider may perceive more continuous care than other women, including those who see two regular providers, we hypothesize that women with one regular provider will report higher satisfaction with care received during the past year. In this sample, 20% of women reported no regular provider, 40% reported one, and 40% reported both a generalist and an obstetrician-gynecologist.

Analytic Methods

Path analysis was used to investigate the strength of causal links between the variables in the three tiers described above. Path analysis involves fitting causal models to the data, which are depicted graphically with diagrams that illustrate causal connections between the variables as straight unidirectional arrows, co-variation between variables as double-headed arrows, and errors as latent random variables. Each of the causal connections is associated with a regression weight. Under a causal model, an estimated variance-covariance matrix can be derived which can be tested for model fit against the sample variance-covariance matrix (Asher 1988; Hatcher 1998; Pedhazur 1982; Wright 1934).

The estimates of the parameters of the models were derived by using a maximum likelihood method. Since many of the variables were not expected to be normally distributed, bootstrapping methods were used to obtain robust standard error estimates and confidence intervals. All analyses were performed on the variance-covariance matrix instead of the correlation matrix. The structural equation modeling program AMOS was used to fit the model and estimate model parameters (Arbuckle and Wothke 1999).

Before proceeding with the analysis, missing data were imputed by using a Markov Chain Monte Carlo Method (Schafer 1997) and using all numerical predictors available. The impact of missing data on the analysis was not expected to be meaningful because only a small proportion of observations (4%) contained missing values.

Since the causal connections between the three tiers described above were not completely specified *a priori*, some exploratory modifications were performed in order to develop plausible models with the best trade-offs in parsimony, substantive interpretation, and goodness of fit. This was achieved by eliminating causal links whose magnitudes were not statistically significant and by inspecting modification indices (Joreskog and Sorbom 1984) in order to add

links that may serve to improve model fit.

Several indices of fit were used to evaluate the models, such as the chi-square statistic, which provides a test for the null hypothesis that the sample covariance matrix stems from the model. Because we have a large sample, the chi-square statistic was expected to be significant; consequently, the null hypothesis that the data fit the covariance matrix of the hypothesized model would be erroneously rejected due to trivial deviations between the fitted model and sample covariance matrices (Hu and Bentler 1995). Therefore, additional indices assessing the congruence between model and data were inspected, such as the normed fit index (NFI) and the comparative fit index (CFI) (Hu and Bentler 1995). The NFI and CFI range from 0 to 1, where 0 represents the fit in the null model in which all variables are modeled as uncorrelated, and 1 represents the fit of the saturated model in which enough parameters exist to replicate the sample covariance matrix without error. According to Bentler and Bonnet (1980), models with a NFI or CFI less than .9 can be substantially improved. Thus, a value of .9 or greater is commonly regarded as indicating goodness of fit.

Because exploratory work was performed in order to find the best fitting models, the above indices may be biased due to the fact that we capitalized on the sample to select the best models. In order to compare the models objectively, it was then necessary to use indices and methods that provide approximate corrections for this bias. The indices used for this purpose adjusted the goodness of fit for model complexity. These included the Akaike Information Criterion, the expected cross-validation index, the minimum discrepancy score divided by the number of degrees of freedom, and the root mean square error of approximation (RMSEA). A ratio of 2 to 3 for the minimum discrepancy score per degrees of freedom indicates an acceptable fit according to Carmines and McIver (1981), and a value of less than .05 indicates a close fit of the model, with .08 or less being reasonable (Cudeck and Browne 1983). A boot-strapping procedure as described in Linhart and Zucchini (1986) and Arbuckle and Wothke (1999) also was performed to compare the fit between models.

After selecting the preferred model, model stability was checked across different groups, including clinic site, age categories, and race/ethnic categories. This was achieved by conducting a multiple-group analysis (Arbuckle and Wothke 1999; Lee and Tsui 1982). First, the model is fit to each subgroup, allowing for the model parameters to vary freely. Second, the regression parameter weights are constrained to be equal across groups. Finally, a likelihood ratio test is conducted between models to judge whether the constrained model does not significantly change the goodness of fit. If the test yields non-significant results, this would imply that there is no heterogeneity in the causal weights across different subgroups. If the test yields significant results, then some of the weights deemed to be heterogeneous are allowed to vary. Models resulting from lesser constraints are compared again until fit is judged not to significantly deteriorate from the unconstrained model.

Because the PCSSW contains both visit-specific and longitudinal components (satisfaction over the past year) and the third tier includes global ratings for both time frames, the causal model was tested on a subset of the sample who had at least one visit to the site prior to

the index visit (n = 1,021).

Results

Table 12 shows the correlation matrix for the variables used in the final path analysis, with means and standard deviations also included. Strong bivariate correlations ranging from .40 to .70 in magnitude exist between the expectations discrepancy score and the endogenous variables of the model. Strong correlations ranging from .42 to .74 are also observed between the PCSSW scales and the two remaining endogenous variables.

A diagram representing the final model is shown in **Figure 6**. Modifications to the model include eliminating the continuity of provider measure because it does not significantly affect any of the endogenous variables and constraining two of the covariances between exogenous variables to be equal to 0 based on the non-significance of the estimates. These constraints were performed in order to preserve degrees of freedom and to reduce model complexity. Specifically, the covariances between site continuity and expectations discrepancy, and between expectations discrepancy and perceived health status, were set to zero. In addition, in order to maintain model parsimony, age was not included in the model. Though age shares a significant but small relationship with the positive rating for the past 12 months as well as the overall visit rating when included (standardized beta = .17 and .09, respectively), its inclusion did not descriptively alter or modify any of the other observed relationships among the substantive variables of interest.

The modification indices, which are estimates of change in model fit when a new parameter is added, suggest that the original model with uncorrelated errors may be misspecified. This means that the PCSSW scales share additional causes in common that are not included in the model. Similarly, the modification indices suggest that the two global outcomes (ratings for care during the past 12 months and at the visit) may have unknown causes in common. In order to further specify the model in the absence of these causes, it was necessary to allow the errors to be correlated, as shown by the double head curves in **Figure 5**. The goodness of fit indices in **Table 13** allow for the comparison between models with uncorrelated and correlated errors. A substantial improvement in the fit can be observed when the errors are correlated.

Goodness of fit indices, which assess how well the model fits the data, for the best fitting model are shown in **Table 13**. The fit indices indicate that this model explains the sample data very well. Although, according to the chi-square test, the model is rejected at this sample size (n = 1,021), the NFI and CFI indices are almost equal to 1, indicating excellent relative fit. The RMSEA (.02) and the Cmin/df (1.2) satisfy the standard criteria for good fit. Furthermore, the AIC and the ECVI of this model are close to the values observed in the saturated model.

Table 14 presents the standardized regression weights and the squared multiple correlations of the best fitting model. The strongest standardized weights above .30 are those corresponding to the weights between expectations discrepancy and the Communication scale

(0.70), expectations discrepancy and the Care Coordination and Comprehensiveness scale (0.63), expectations discrepancy and the Administration and Office Procedures scale (0.64), Care Coordination and Comprehensiveness and the CAHPS rating (0.54), and Communication and the visit rating (0.46). Weaker standardized significant weights below .30 include those between Care Coordination and Comprehensiveness and the visit rating (0.25), Administration and Office Procedures and the visit rating (0.13), Communication and the CAHPS rating (0.09), site continuity and Care Coordination and Comprehensiveness (0.04), and perceived health status and Communication (0.06). The squared multiple correlations indicate that the Communication scale has approximately 49% of its variance explained by its predictors, and Care Coordination and Comprehensiveness has 40% of its variance explained. The CAHPS rating, Administration and Office Procedures, and the visit rating have 38%, 41%, and 60% of their variances explained, respectively.

Using multiple-group analysis, no statistically significant differences among regression weights are observed when the model is fitted separately across different race/ethnic subgroups (white, black, and other) or age groups (18-35, 36-54, and 55 and over). However, some weights had to be allowed to vary freely between the three different sites in order for the overall model not to be rejected by the likelihood ratio test. These weights were those assessing the relationship between expectations discrepancy and Care Coordination and Comprehensiveness (.54, .67, .68 for sites 1, 2, and 3, respectively) as well as those between Care Coordination and Comprehensiveness and the CAHPS rating (.58, .41, .35, respectively). These variations suggest that the first site shows significantly modified strengths of the relationship among these variables compared to the other sites. In order to illustrate the degree of model stability across sites, stable links in **Figure 6** are drawn in thicker print.

There are several limitations to the path analysis. Except for face validity, the validity of the some of the key variables in the model such as the expectations discrepancy score and the overall visit rating may be subjected to scrutiny. Not accounting for the actual reliabilities of the variables may affect the estimates in several ways. If the causal variables have low reliability, the standardized regression weights may be meaningfully attenuated, leading to underestimation of the strength in the relationships (Pedhazur 1982). Less than perfect reliabilities on the dependent variables will also result in under-estimation of the proportions of variance explained. In any event, the magnitudes presented in this analysis are lower bounds of magnitudes that would occur if the variables were to be perfectly reliable.

The non-normality of the variables may also bias the parameter estimates, their standard errors, hypothesis tests, and indices of fit (West et al. 1995). Coarsely categorized variables, for instance, may lead to further attenuation of the relationships. However, the adverse effect of variable skewness on our inferences is minimized because of the large sample size ($n = 1,021$) and our use of boot-strapping methods to calculate robust standard errors.

Overall, the results of the path analysis largely affirm our theoretical model. The results show that perceived health status and site continuity have small but significant indirect effects on global ratings of care by way of satisfaction with Communication and with Care Coordination

and Comprehensiveness. Expectations discrepancies scores are strongly associated with each PCSSW scale. This result is important because it confirms that expectations are related to satisfaction and have an indirect path to the global ratings of quality of care. Further, the PCSSW scales are differentially related to the global ratings in the predicted pattern: Care Coordination and Comprehensiveness is most highly correlated with the CAHPS rating of care over the past 12 months, whereas Communication has the highest correlation with the global visit rating. Administration and Office Procedures contributes only to the visit rating. These findings suggest that visit-based ratings of care and longitudinal ratings of care, though often used interchangeably in the literature, are reflecting distinct aspects of health care experiences.

DISCUSSION

Main Findings and Implications

The PCSSW is a 24-item survey tool consisting of three scales that have been shown to be psychometrically reliable and valid. The PCSSW may be used in either self-administered or telephone format, and it represents an important new approach for measuring women's satisfaction levels with their primary care in both traditional and women-specific health settings. The PCSSW should be useful for quality improvement efforts within health plans or health care programs.

The PCSSW contains three discrete and interpretable scales: Communication at the visit (8 items); Administration and Office Procedures at the visit (6 items); and Care Coordination and Comprehensiveness over the past 12 months (10 items). Developed through careful qualitative work involving women with diverse social and health characteristics across the country, the PCSSW has both visit-specific and longitudinal components; demonstrates factor invariance across subgroups of women defined by age group and by race/ethnicity; strong psychometric properties in the field test; and added explanatory power compared to a generic satisfaction measure.

Path analysis of the determinants of satisfaction and global ratings of quality of care shows that the PCSSW scales articulate with both visit-specific and longitudinal aspects of health care experiences. Ratings of visit quality were influenced primarily by the PCSSW Communication scale, whereas ratings of quality of care for the past year were influenced mainly by the PCSSW Care Coordination and Comprehensiveness scale. These findings demonstrate the depth of the PCSSW in capturing both short- and longer-term care experiences and the dimensions of health care relevant to how patients value their care. The results also show evidence that global ratings of quality of care have different meanings: day-of-visit ratings are more reflective of communication with the health care provider than are longitudinal ratings; the latter are more reflective of ongoing health care management experiences. For either global rating of quality of care (day-of-visit or past year), patient expectations have a key role in how patients cognitively summarize their satisfaction with care. Patients whose expectations were met or exceeded were more highly satisfied with their care than those whose expectations were not met (as determined by the discrepancy between pre-visit goals for the visit and post-visit

reports of what occurred).

The content of the PCSSW differs from existing satisfaction tools in several ways. Some of the items are specific to women (e.g., the ability to get both gynecological and general health care at the same site; the health professional's knowledge of women's health issues). Other items are new topics not typically included in patient satisfaction surveys but potentially applicable to all patients (e.g., how well office staff keeps the patient informed about waiting time; the health professional's interest in the patient's mental and emotional health). Additional items are similar to items in generic patient satisfaction surveys but are worded to be consistent with women's framings as discovered in the focus groups (e.g., the health professional's ability to answer questions in a "sensitive and caring way"). The items also address both a specific visit and care coordination and comprehensiveness during the past year. Because many women seek health care from more than one professional or site, the latter component of the PCSSW is particularly innovative.

The Care Coordination and Comprehensiveness scale contains the most new content and also demonstrates consistently high performance across the validity assessments. The Communication scale also performs well. Both of the visit-specific scales, the Communication scale and the Administration and Office Procedures scale, are more like the interpersonal and technical aspects of quality tapped in generic tools such as the MOS Visit Satisfaction scale. The PCSSW adds sensitivity to satisfaction measurement and can be useful in evaluations of the quality of primary care and in quality improvement programs.

Applications

To date, the PCSSW has been used in two studies assessing patient satisfaction in different organizational settings for women's primary care. First, in a 10-site survey of women veterans served in Veterans Administration (VA) women's clinics compared with traditional VA primary care clinics, the draft version of the PCSSW was used to provide satisfaction scores in five domains identified in the focus groups: getting care; privacy and comfort; communication; complete care; and follow-up care. In adjusted comparisons, women served in VA women's clinics reported significantly higher satisfaction on all five domains (Bean-Mayberry et al. 2003).

Second, in the national evaluation of the quality of care in the clinical care centers of the National Centers of Excellence in Women's Health (CoEs), the PCSSW was administered in a telephone survey to test the hypothesis that women served in CoEs report higher satisfaction than women in community comparison samples. Only the Care Coordination and Comprehensiveness scale was used in these comparisons because the surveys measured satisfaction with care in the past year rather than with a specific health care visit. In adjusted comparisons, women who were patients in the CoEs reported significantly higher satisfaction on this scale than women in the community, with a strong effect size (Cohen's $d = 0.449$) (Anderson et al. 2002).

Both of these projects demonstrate that the PCSSW is useful for evaluating patient satisfaction in diverse primary care settings, including programs designed for women.

Study Limitations

This project was limited to cross-sectional data, so associations between satisfaction and subsequent behaviors (e.g., returning to the same provider or site, following the provider's recommendations) could not be observed. The study sample also was limited to patients served in settings affiliated with three health systems, although two additional studies (described above) have used the PCSSW to date. Use of the PCSSW in other locations and samples is recommended. Finally, the results of this project, which demonstrate the usefulness of the PCSSW, provide a basis for translating the tool and testing it in other languages.

Conclusion

Outpatient satisfaction tools designed to be used in general populations fail to capture the full range of health care concerns of women, as demonstrated by our qualitative research. The PCSSW is a psychometrically valid survey tool for assessing women's satisfaction with primary care, and it has been shown to have greater explanatory power compared to a generic instrument. It may be self-administered or conducted by telephone interview. The PCSSW may be used both in studies to evaluate alternative models of primary care delivery for women and in quality improvement programs in women's primary care.

PROJECT REPORTS AND PUBLICATIONS

PCSSW Users' Manual

The PCSSW and user's Manual, **Primary Care Satisfaction Survey for Women: A Manual for Use and Scoring**, may be obtained from Richard D. Lennox, Ph.D., Psychometric Technologies Inc., 402 Millstone Drive, Hillsborough, NC 27279 (Phone: 919-245-0930; email: Rlennox@mindspring.com).

Publications

Anderson, Roger T., Carol S. Weisman, Sarah Hudson Scholle, Jillian T. Henderson, Robert Oldendick, and Fabian Camacho. 2002. "Evaluation of the Quality of Care in the Clinical Care Centers of the National Centers of Excellence in Women's Health." *Women's Health Issues* 12(6):309-326.

Working Papers

Scholle, Sarah Hudson, Carol S. Weisman, Roger T. Anderson, and Fabian Camacho. "The Development and Validation of the Primary Care Satisfaction Survey for Women (PCSSW)"

Anderson, Roger T., Carol S. Weisman, Sarah Hudson Scholle, and Fabian Camacho. "Predictors of Women's Satisfaction with their Primary Care Using Two Different Rating Approaches."

Dissertation in Progress

Henderson, Jillian T. "Correlates of Women's Primary Care Satisfaction: Experiences and Expectations Across the Lifespan." Ph.D. Dissertation in the Department of Health Management and Policy, Program in Health Services Organization and Policy, University of Michigan School of Public Health, Ann Arbor, Michigan.

REFERENCES

- Aharony L, and Strasser S. 1993. "Patient Satisfaction: What We Know About and What We Still Need to Explore." Medical Care Review 50:49-79.
- Anderson RT, Barbara AM, Weisman C, Scholle SH, Binko J, Schneider T, Freund K, and Gwinner V. 2001. "A Qualitative Analysis of Women's Satisfaction with Primary Care from a Panel of Focus Groups in the National Centers of Excellence in Women's Health." Journal of Women's Health and Gender-Based Medicine 10(7):637-647.
- Anderson RT, Weisman CS, Scholle SH, Henderson JT, Oldendick R, and Camacho F. 2002. "Evaluation of the Quality of Care in the Clinical Care Centers of the National Centers of Excellence in Women's Health." Women's Health Issues 12(6):309-326.
- Arbuckle JL, and Wothke W. 1999. Amos 4.0 User's Guide. Chicago, IL: Smallwaters Corporation.
- Asher HB. 1988. Causal Modeling, 2nd Edition. Beverly Hills, CA: Sage.
- Bandura A, Adams NE, and Beyer J. 1977. "Cognitive Processes Mediating Behavior Change." Journal of Personality and Social Psychology 35:125-139.
- Bean-Mayberry BA, Chang CH, McNeil MA, Whittle JC, Hayes PM, and Scholle SH. 2003 "Patient Satisfaction in Women's Clinics and Traditional Primary Care Clinics in the Veterans Administration." Journal of General Internal Medicine 18:175-181.
- Becker LA. 1999. "Measures of Effect Size (Strength of Association)." November 8, 1999. Available at: http://web.uccs.edu/lbecker/SPSS/glm_effectsize.htm. Accessed August, 2002.
- Bell RA, Kravitz RL, Thom D, et al. 2001. "Unsaid But Not Forgotten: Patients' Unvoiced Desires in Office Visits." Archives of Internal Medicine 161:1977-1984.
- Bentler PM, and Bonnet DG. 1980. "Significance Tests and Goodness of Fit in the Analysis of Covariance Structures." Psychological Bulletin 88:588-606.
- Browne MW, and Cudeck R. 1993. "Alternative Ways of Assessing Model Fit." Pp. 136-162 in KA Bollen and JS Long (eds.), Testing Structural Equation Models. Newbury Park, CA: Sage.
- Carmines EG, and McIver JP. 1981. "Analyzing Models with Unobserved Variables." In GW Bohrnstedt and EF Borgatta (eds.). Social Measurement: Current Issues. Beverly Hills, CA: Sage.
- Clancy CM, and Massion CT. 1992. "American Women's Health Care: A Patchwork Quilt with Gaps." Journal of the American Medical Association 268:1918-1920.
- Cleary PD, and McNeil BJ. 1988. "Patient Satisfaction as an Indicator of Quality Care." Inquiry 25:25-36.
- Cudeck R, and Browne MW. 1983. "Cross-validation of Covariance Structures." Multivariate Behavioral Research 18:147-167.

- Davies AR, and Ware JE. 1991. GHAA's Consumer Satisfaction Survey and User's Manual. Washington, DC: Group Health Association of America.
- Donaldson MS. 2001. "Continuity of Care: A Reconceptualization." Medical Care Research and Review 58(3):255-290.
- Gallagher TC, Geling O, and Comite F. 2001. "Use of Multiple Providers for Regular Care and Women's Receipt of Hormone Replacement Therapy Counseling." Medical Care 39(10):1086-1096.
- Gill JM, and Mainous AG. 1998. "The Role of Provider Continuity in Preventing Hospitalizations." Archives of Family Medicine 7:352-357.
- Hall JA, and Dornan MC. 1990. "Patient Sociodemographic Characteristics as Predictors of Satisfaction with Medical Care: A Meta-Analysis." Social Science and Medicine 30(7):811-818.
- Hatcher L. 1998. A Step-by-step Approach to Using the SAS System for Factor Analysis and Structural Equation Modeling. Cary, NC: SAS Institute.
- Hays RD, Shaul JA, Williams VSL, et al. 1999. "Psychometric Properties of the CAHPS™ 1.0 Survey Measures." Medical Care 37(3) supplement: MS22-MS31.
- Henderson JT, and Weisman CS. 2001. "Physician Gender Effects on Preventive Screening and Counseling: An Analysis of Male and Female Patients' Health Care Experiences." Medical Care 39(12):1281-1291.
- Henderson JT, Weisman CS, and Grason H. 2002. "Are Two Doctors Better Than One? Women's Physician Use and Appropriate Care." Women's Health Issues 12(3):138-149.
- Horn JL. 1965. "A Rationale and Test for the Number of Factors in Factor Analysis." Psychometrika 30:179-185.
- Hu L, and Bentler PM. 1995. "Evaluating Model Fit." In RH Hoyle (ed.), Structural Equation Modeling: Concepts, Issues, and Applications. Thousand Oaks, CA: Sage.
- Joreskog KG, and Sorbom D. 1984. Lisrel-vi User's Guide, 3rd edition. Mooresville, IN: Scientific Software.
- Kaiser HF. 1970. "A Second Generation Little Jiffy." Psychometrika. 35:401-415.
- Kolodinsky J. 1997. "Gender Differences in Satisfaction with Primary Care Physicians in a Managed Care Health Plan." Women and Health 26:67-86.
- Kravitz RL, Callahan EJ, Paterniti D, et al. 1996. "Prevalence and Sources of Patients' Unmet Expectations for Care." Annals of Internal Medicine 124(9):730-737.
- Lee SY, and Tsui KL. 1982. "Covariance Structure Analysis in Several Populations." Psychometrika 47:297-308.

- Linhart H, and Zucchini W. 1986. Model Selection. New York: Wiley.
- Magill M, and Senf J. 1987. "A New Method for Measuring Continuity of Care in Family Practice Residencies." Journal of Family Practice 24:165-168.
- National Committee for Quality Assurance. 1998. HEDIS 1999 Volume 3: HEDIS Protocol for Administering CAHPS 2.0H Survey. Washington, D.C.: NCQA.
- O'Malley AS, Forrest CB, and O'Malley PG. 2000. "Low-income Women's Priorities for Primary Care." Journal of Family Practice 49(2):141-146.
- Orlando M, and Meredith LS. 2002. "Understanding the Causal Relationship Between Patient-reported Interpersonal and Technical Quality of Care for Depression." Medical Care 40(8):696-704.
- Peck BM, Asch DA, Goold SD, et al. 2001. "Measuring Patient Expectations: Does the Instrument Affect Satisfaction or Expectations?" Medical Care 39(1):100-108.
- Pedhazur EJ. 1982. Multiple Regression in Behavioral Research, 2nd Edition. New York: Holt.
- Rubin HR, Gandek B, Rogers WH, et al. 1993. "Patient's Ratings of Outpatient Visits in Different Practice Settings: Results from the Medical Outcomes Study." Journal of the American Medical Association 270(7):835-840.
- Safran DG, Kosinski M, Tarlov AR, et al. 1998. "The Primary Care Assessment Survey: Tests of Data Quality and Measurement Performance." Medical Care 36(5):728-739.
- Safran DG, Tarlov AR, and Rogers WH. 1994. "Primary Care Performance in Fee-for-Service and Prepaid Health Care Systems: Results from the Medical Outcomes Study." Journal of the American Medical Association 271(20):1579-1586.
- Samejima F. 1997. "Models for Items with Polytomous Response Formats: Graded Response Model." In WJ Linden and RK Hambleton (eds.), Handbook of Modern Item Response Theory. New York: Springer-Verlag.
- Schafer JL. 1997. Analysis of Incomplete Multivariate Data. New York: Chapman and Hall.
- Schauffler HH, Rodriguez T, and Milstein A. 1996. "Health Education and Patient Satisfaction." Journal of Family Practice 42(1):62-68.
- Scholle SH, Weisman CS, Anderson R, Weitz T, Freund KM, and Binko J. 2000. "Women's Satisfaction with Primary Care: A New Measurement Effort from the PHS National Centers of Excellence in Women's Health." Women's Health Issues 10(1):1-9.
- Sitzia J, and Wood N. 1997. "Patient Satisfaction: A Review of Issues and Concepts." Social Science and

Medicine 45(12):1829-1843.

Velicer WF. 1976. "The Relation Between Factor Score Estimates, Image Scores, and Principal Component Scores." Educational and Psychological Measurement 36:149-159.

Ware JE, and Sherbourne CD. 1992. "The MOS 36-Item Short-Form Health Survey (SF-36). I. Conceptual Framework and Item Selection." Medical Care 30(6):473-483.

Weisman CS. 1996. "Women's Use of Health Care," Chapter 1 in MM Falik and KS Collins (eds.), Women's Health: The Commonwealth Fund Survey. Baltimore, MD: Johns Hopkins University Press.

Weisman, CS, and Henderson JT. 2001. "Managed Care and Women's Health: Access, Preventive Services, and Satisfaction." Women's Health Issues 11(3):201-215.

Weisman CS, Henderson JT, Schiffrin E, Romans M, and Clancy CM. 2001. "Gender and Patient Satisfaction in Managed Care Plans: Analysis of the 1999 HEDIS/CAHPS 2.0H Adult Survey." Women's Health Issues 11(5):401-415.

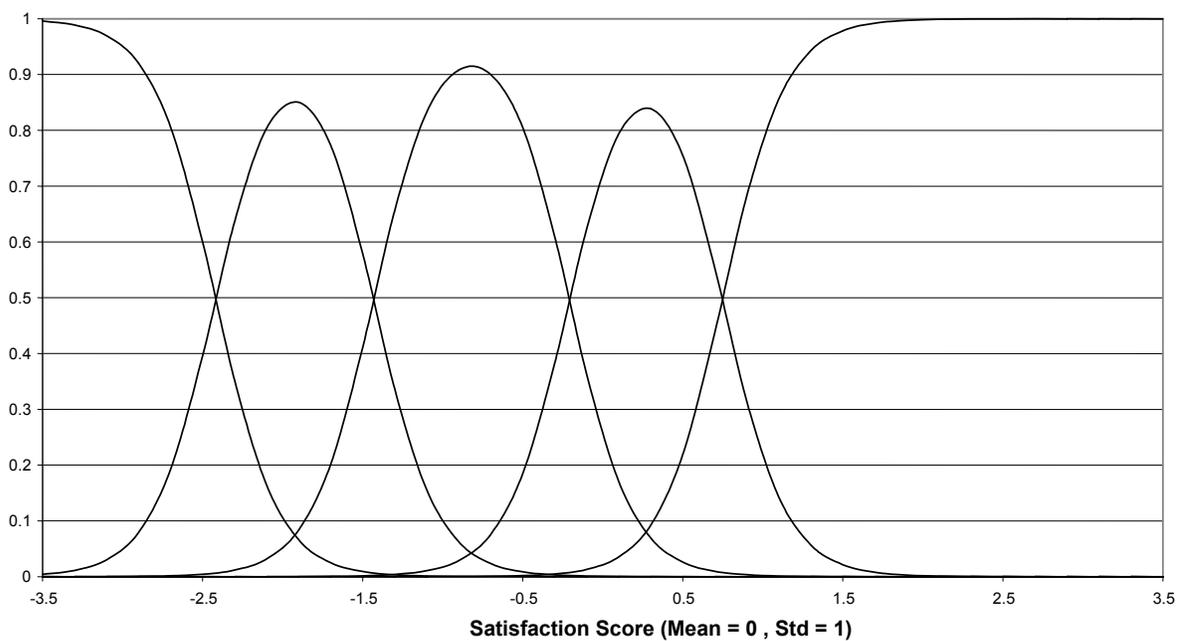
Weisman CS, Rich DE, Rogers J, et al. 2000. "Gender and Patient Satisfaction with Primary Care: Tuning in to Women in Quality Measurement." Journal of Women's Health and Gender-based Medicine 9(6):657-665.

West GW, Finch JF, and Curran PJ. 1995. "Structural Equation Models with Nonnormal Variables." In RH Hoyle (ed.), Structural Equation Modeling: Concepts, Issues, and Applications. Thousand Oaks, CA: Sage.

Wright S. 1934, "The Method of Path Coefficients." Annals of Mathematical Statistics 5:161-215.

Probability

Figure 1. Samejima Item Characteristic Curves for Item Q11i



Trace lines showing the probability of selecting the lowest to highest response at a given level of satisfaction are displayed in order

Figure 2. Information Curves for Communication Items

I(theta)

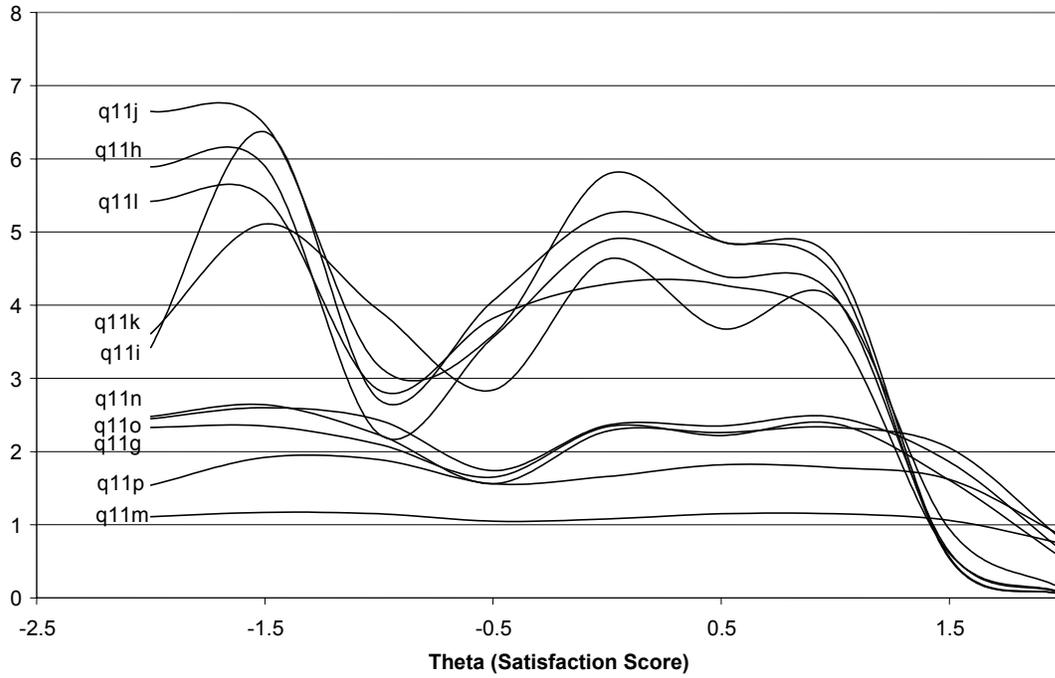


Figure 3. Information Curves for Administration and Office Procedures Items

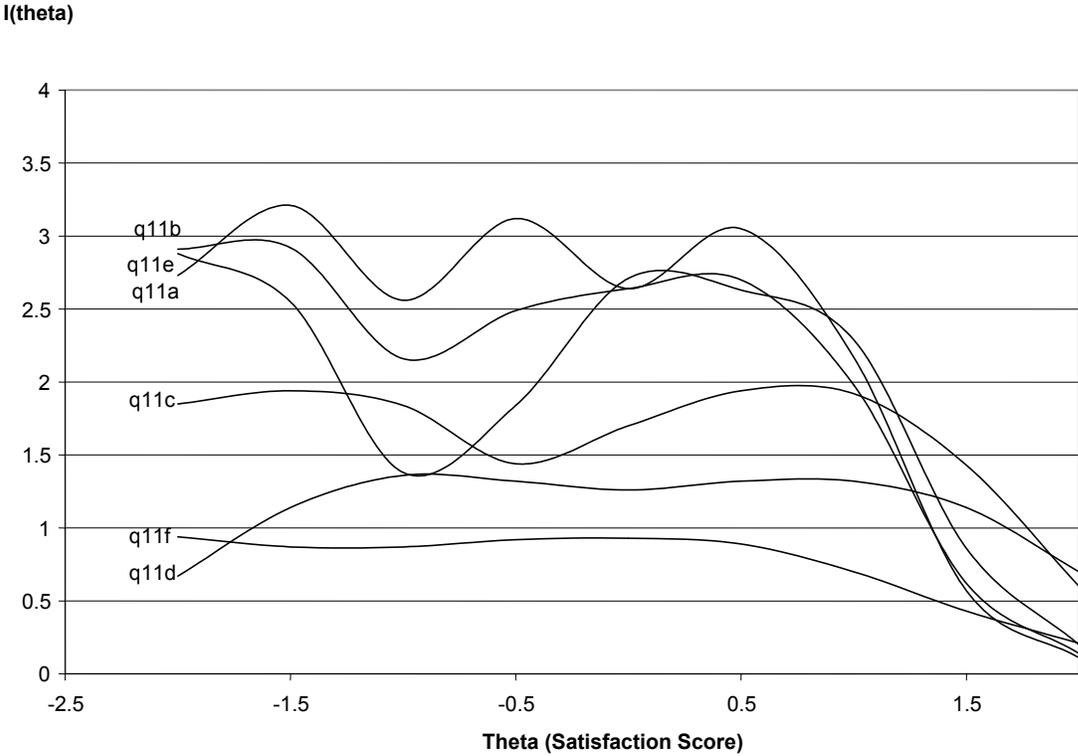


Figure 4. Information Curves for Care Coordination and Comprehensiveness Items

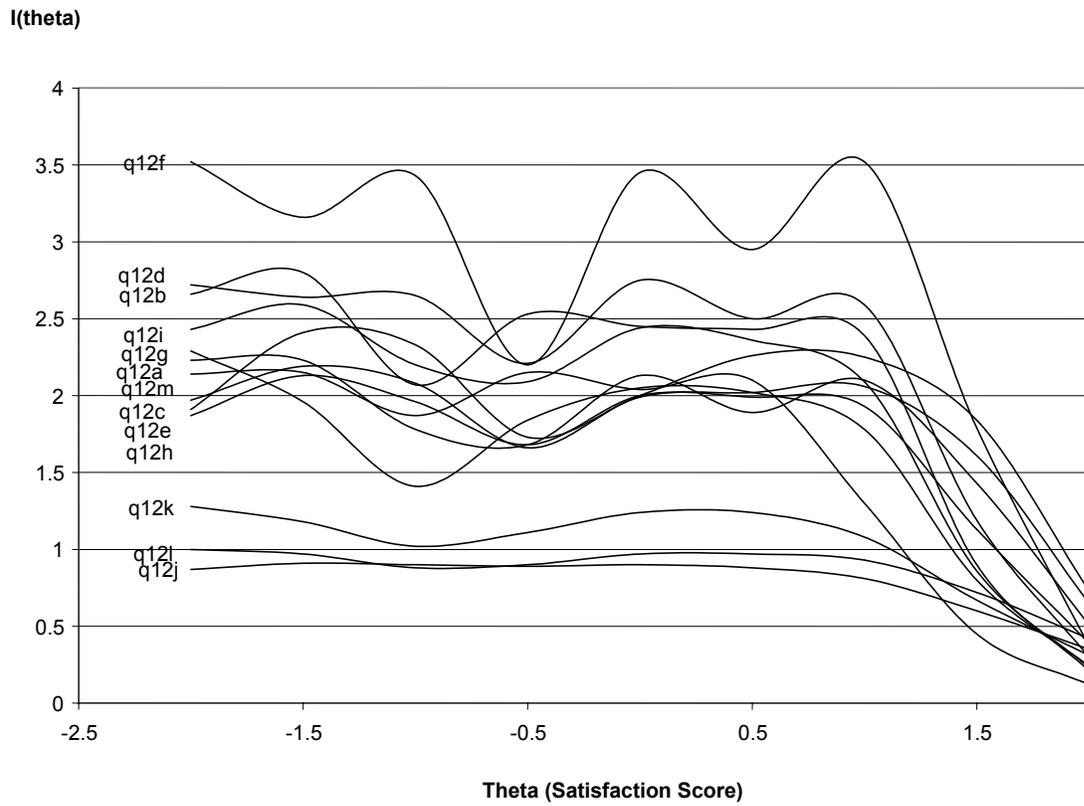


Figure 5. Structural Equation Model to Test Invariance Across Age and Race/ethnic Groups (n = 1,202)

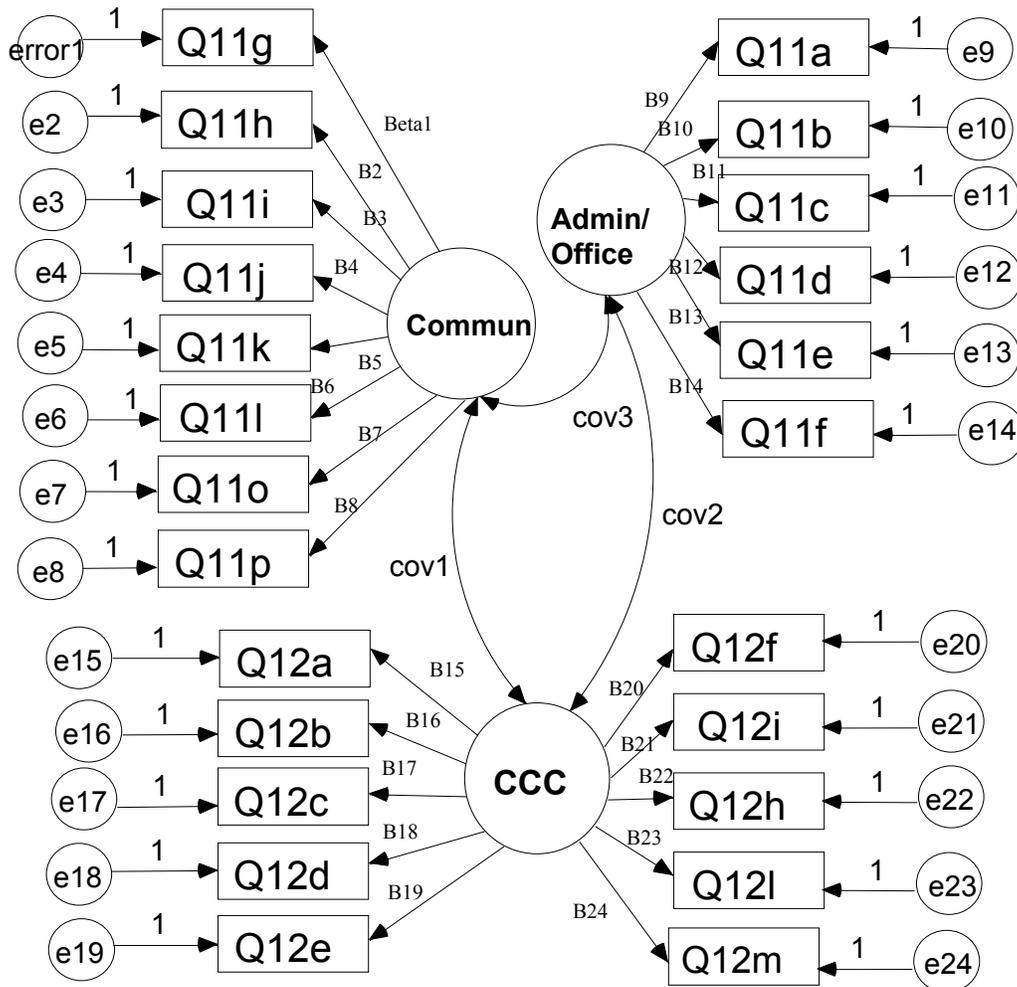
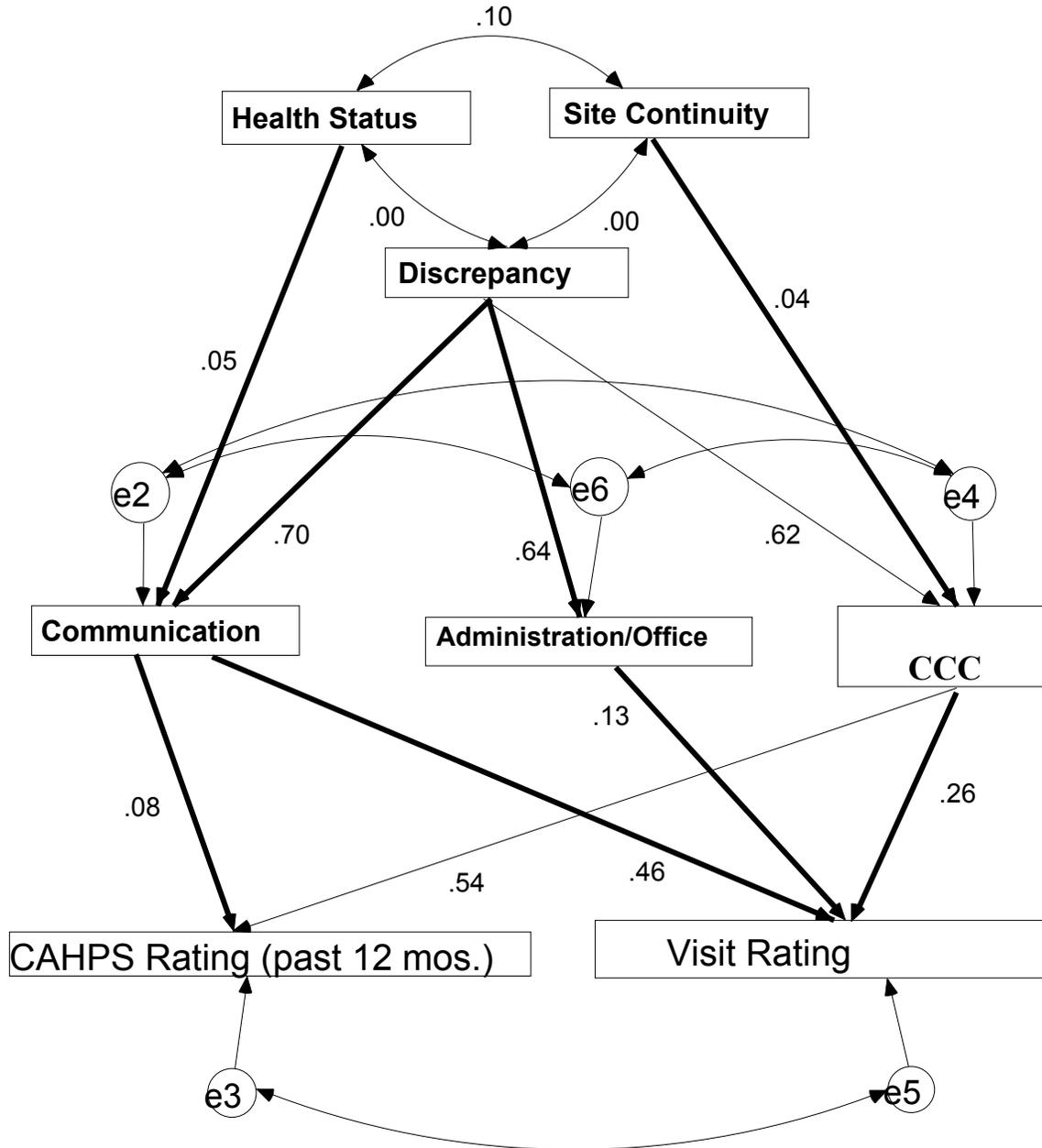


Figure 6. Path Analysis Model (n = 1,021)^a



^a Significant standardized regression weights and correlations are shown. Thick arrows represent magnitudes that are stable across sites.

Table 1. Characteristics of the Pooled Study Sample (n = 1,202)

	% or Mean (Standard Deviation)
<u>Demographic and Health Characteristics</u>	
Age (mean, SD)	41.81 (16.56)
Education (%):	
High school or less	30.35
Some college	31.28
College graduate	18.80
Graduate school	19.56
Household income (%):	
\$20,000 or less	25.95
\$20,001 - \$50,000	30.71
\$50,001 - \$75,000	15.47
\$75,001 and over	16.97
Refused/Don't know	0.90
Health insurance ^a (%):	
Private	62.52
Medicaid	18.86
Medicare	11.52
Other/None	7.10
Race/ethnicity (%):	
White, non-Hispanic	66.53
Black, non-Hispanic	23.48
Other/Multi-ethnic ^b	9.99
Perceived health status ^c (%):	
Excellent	13.42
Very Good	35.00
Good	34.33
Fair	14.50
Poor	2.75
Pregnant in past year (%)	18.58

Table 1. Characteristics of the Pooled Study Sample (continued)

Health Care Utilization

Number of office visits, past year (mean, SD)	7.79 (8.24)
Length of time coming to clinical site (%):	
First time today	10.57
Less than one year	20.80
1 –2 years	16.31
2 years or more	52.33
Clinical site is usual source of care (%)	76.44
Main reason for visit (%):	
Followup care	31.65
New health problem	26.53
Routine exam	25.78
Prenatal or postpartum care	16.04
Type(s) of regular provider (%):	
Generalist physician + Obgyn	40.20
Generalist physician	25.93
Obgyn	10.05
Other health professional	3.89
No regular provider	19.93
Regular provider (for those who have one) is at this site (%)	78.25

a “Other” health insurance includes military, CHAMPUS, Tri Care, or the VA. Respondents reporting more than one source of health insurance were coded hierarchically in the following order: Medicaid, Medicare, private, other, none.

b “Other” race/ethnicity includes those with Hispanic ancestry, other racial/ethnic identification, and multi-ethnic identification.

c This item is from the SF-36: “In general, would you say your health is...” (Ware and Scherbourne 1992). When used as a covariate in analyses, it is dichotomized to contrast those reporting “fair” or “poor” health with all others.

Table 2. Exploratory Factor Loadings (n = 601)*

First Item Set: Today's Visit		Loadings Factor 1	Loadings Factor 2
<i>Items loading on Factor 1 (Communication)</i>			
Q11i	My health professional's ability to explain things clearly	0.97	-0.06
Q11j	My health professional's ability to help me feel comfortable talking about my concerns	0.95	-0.02
Q11l	My health professional's ability to take what I say seriously	0.94	-0.01
Q11k	The chance to ask all my questions	0.92	0.01
Q11h	My health professional's ability to answer questions in a sensitive and caring way	0.92	0.02
Q11o	My health professional's willingness to explain different options for my care	0.78	0.11
Q11p	My health professional's interest in how my life affects my health	0.76	0.06
Q11g	The amount of time I had to talk with my health professional	0.67	0.25
Q11n	The chance to get everything I need at this visit	0.72	0.20
Q11m	My health professional's knowledge of my medical history	0.59	0.17
<i>Items loading on Factor 2 (Administration & Office Procedures)</i>			
Q11a	The courtesy of the office staff	0.10	0.76
Q11b	The staff's flexibility in scheduling my appointment around my needs	0.08	0.78
Q11c	Privacy when talking to the receptionist	0.06	0.66
Q11d	How well the staff kept me informed about the waiting time	-0.03	0.72
Q11e	Help with scheduling my next visit	0.07	0.75
<i>Item loading both factors</i>			
Q11f	The chance to talk to my health professional with my clothes on	0.36	0.40
<i>% variance explained by both factors</i>			70%
<i>Inter-factor correlation</i>		.61	

Table 2. Exploratory Factor Loadings (continued)

Second Item Set: Care During the Past 12 Months		Loadings Factor 1
Q12f	How well my health care fits my stage of life	0.87
Q12d	The health professionals' interest in my mental and emotional health	0.86
Q12c	The information I get about healthy living (such as diet and exercise)	0.83
Q12i	How well the health professionals explain the results of tests or procedures	0.83
Q12a	The health professional's focus on prevention	0.82
Q12b	The health professional's knowledge of women's health issues	0.82
Q12e	Help with finding information resources in women's health	0.81
Q12l	The chance to get both gynecological and general health care here	0.81
Q12h	Information about how to get the results of my tests	0.80
Q12m	My overall trust in the health professionals here	0.69
Q12g	How well my health information is kept private	0.79
Q12k	The chance to see a health professional of the gender I prefer	0.76
Q12j	The chance to see the same health professional at each visit	0.64
<i>% variance explained by factor</i>		64%

* Results of a principal factor analysis with a Promax rotation

Table 3. Item Response Theory Results: Samejima Graded Model Item Parameters and Standard Errors

Abbreviated Item Content	a	b ₁	b ₂	b ₃	b ₄
Today's Visit Items					
Factor 1: Communication					
Q11g: Amount of time to talk	3.01 (0.14)	-2.25 (0.17)	-1.27 (0.13)	0.09 (0.05)	1.10 (0.04)
Q11i: Explain things clearly	5.09 (0.28)	-2.42 (0.18)	-1.43 (0.06)	-0.21 (0.05)	0.75 (0.03)
Q11j: Help me feel comfortable talking	5.13 (0.28)	-2.08 (0.15)	-1.38 (0.08)	-0.16 (0.04)	0.76 (0.03)
Q11k: Chance to ask all of my questions	4.29 (0.19)	-1.76 (0.10)	-1.22 (0.06)	-0.09 (0.05)	0.83 (0.04)
Q11l: Take what I say seriously	4.52 (0.24)	-2.02 (0.18)	-1.41 (0.07)	-0.23 (0.04)	0.74 (0.03)
Q11h: Sensitive and caring answers	4.73 (0.27)	-2.10 (0.28)	-1.49 (0.08)	-0.19 (0.04)	0.76 (0.03)
Q11p: Interest in how my life affects my health	2.58 (0.14)	-1.74 (0.14)	-0.97 (0.09)	0.31 (0.05)	1.32 (0.05)
Q11o: Willingness to explain different options	3.08 (0.15)	-1.92 (0.16)	-1.14 (0.09)	0.17 (0.04)	1.17 (0.04)
Q11m: Knowledge of my medical history	2.01 (0.11)	-2.02 (0.17)	-1.04 (0.08)	0.36 (0.05)	1.42 (0.06)
Q11n: The chance to get everything I need	3.07 (0.16)	-1.95 (0.13)	-1.23 (0.09)	0.18 (0.05)	1.25 (0.04)
Factor 2: Administration and Office Procedures					
Q11a: Courtesy of the office staff	3.50 (0.21)	-2.41 (0.21)	-1.43 (0.07)	-0.38 (0.04)	0.65 (0.04)
Q11b: Flexibility in scheduling my appointment	3.22 (0.18)	-2.09 (0.15)	-1.44 (0.08)	-0.26 (0.04)	0.67 (0.04)
Q11c: Privacy when talking to the receptionist	2.63 (0.14)	-1.98 (0.14)	-1.11 (0.07)	0.28 (0.05)	1.15 (0.06)
Q11d: Informed about the waiting time	2.12 (0.12)	-1.29 (0.09)	-0.67 (0.06)	0.44 (0.05)	1.32 (0.07)
Q11e: Help with scheduling next visit	3.18 (0.18)	-2.02 (0.14)	-1.66 (0.10)	-0.09 (0.04)	0.78 (0.05)
Q11f: Talk with my clothes on	1.79 (0.12)	-2.60 (0.24)	-2.02 (0.16)	-0.55 (0.07)	0.56 (0.06)

Table 3. Item Response Theory Results: Samejima Graded Model Item Parameters and Standard Errors (continued)

Abbreviated Item Content	a	b ₁	b ₂	b ₃	b ₄
Care During Past Year Items					
Q12a: Focus on prevention	2.83 (0.13)	-2.20(0.15)	-1.38 (0.08)	-0.03 (0.04)	1.08 (0.05)
Q12b: Knowledge of women's health issues	3.24 (0.16)	-2.36 (0.19)	-1.49 (0.09)	-0.29 (0.04)	0.80 (0.04)
Q12c: Information about healthy living	2.78 (0.13)	-1.88 (0.10)	-1.09 (0.06)	0.18 (0.04)	1.18 (0.05)
Q12d: Interest in my mental and emotional health	3.21 (0.16)	-2.02 (0.12)	-1.14 (0.07)	-0.03 (0.04)	0.92 (0.04)
Q12e: Information resources in women's health	2.91 (0.14)	-1.75 (0.10)	-1.02 (0.07)	0.28 (0.04)	1.23 (0.05)
Q12f: Care fits my stage of life	3.70 (0.17)	-1.98 (0.12)	-1.11 (0.06)	0.09 (0.04)	1.04 (0.04)
Q12g: Information is kept private	2.78 (0.13)	-2.28 (0.17)	-1.70 (0.10)	-0.24 (0.05)	0.77 (0.05)
Q12h: Get the results of my tests	2.71 (0.13)	-1.87 (0.12)	-1.16 (0.20)	0.05 (0.05)	0.96 (0.05)
Q12i: Explain results of tests or procedures	3.01 (0.14)	-1.98 (0.12)	-1.28 (0.07)	-0.13 (0.04)	0.79 (0.05)
Q12j: Same health professional at each visit	1.73 (0.10)	-2.11 (0.16)	-1.25 (0.09)	-0.13 (0.06)	0.92 (0.07)
Q12k: Health professional of the gender I prefer	2.08 (0.11)	-2.47 (0.18)	-1.67 (0.11)	-0.14 (0.05)	0.81 (0.06)
Q12l: Both gynecological and general health care	1.84 (0.10)	-2.46 (0.20)	-1.58 (0.11)	-0.07 (0.06)	1.03 (0.07)
Q12m: Overall trust in health professionals	2.82 (0.15)	-2.48 (0.20)	-1.59 (0.09)	-0.46 (0.05)	0.52 (0.04)

a = discrimination index (see text)

b₁, ..., b₄ = item difficulty parameters (see text)

Table 4. Factor Analysis using Maximum Likelihood Factor Analysis on both the Initial and Test Data Sets

	Loadings Factor 1		Loadings Factors 2	
First Item Set: Today's Visit	Standardized Regression Coefficients			
Factor 1: Communication	INITIAL N= 601	TEST N= 601	INITIAL N= 601	TEST N = 601
Q11j: Help me feel comfortable talking	0.97	0.97	-0.02	-0.03
Q11i: Explain things clearly	0.98	0.92	-0.04	0.02
Q11l: Take what I say seriously	0.94	0.90	0.01	0.02
Q11h: Sensitive and caring answers	0.93	0.90	0.03	0.05
Q11k: Chance to ask all of my questions	0.92	0.89	0.03	-0.02
Q11o: Willingness to explain different options	0.70	0.76	0.17	0.11
Q11g: Amount of time to talk	0.67	0.72	0.26	0.17
Q11p: Interest in how my life affects my health	0.68	0.68	0.11	0.14
Factor 2: Administration and Office Procedures				
Q11c: Privacy when talking to the receptionist	0.06	-0.04	0.64	0.83
Q11e: Help with scheduling next visit	0.05	0.09	0.78	0.76
Q11b: Flexibility in scheduling my appointment	0.08	0.08	0.80	0.72
Q11d: Informed about the waiting time	-0.02	-0.02	0.69	0.72
Q11a: Courtesy of the office staff	0.08	0.20	0.79	0.66
Q11f: Talk with my clothes on	0.33	0.36	0.42	0.40
<i>% Variance Explained by Both Factors</i>	INITIAL = 71% ,TEST = 70%			
<i>Interfactor Correlation</i>	INITIAL = .60, TEST = .60			
	Loadings Factor1			
Second Item Set	Standardized Regression Coefficients			
	INITIAL N=601	TEST N= 601		
Q12f: Care fits my stage of life	0.88	0.89		
Q12d: Interest in my mental and emotional health	0.88	0.85		
Q12b: Knowledge of women's health issues	0.83	0.85		
Q12e: Information resources in women's health	0.83	0.83		
Q12i: Explain results of tests or procedures	0.81	0.82		
Q12a: Focus on prevention	0.83	0.82		
Q12c: Information about healthy living	0.85	0.82		
Q12m: Overall trust in health professionals	0.78	0.79		
Q12h: Get the results of my tests	0.77	0.79		
Q12l: Both gynecological and general health care	0.65	0.66		
<i>% Variance Explained by Factor</i>	INITIAL = 66%, TEST = 66%			

Table 5. Primary Care Satisfaction Survey for Women: Scales and Item Univariate Statistics*
(n = 1,202)

Scale and item	Mean (Standard Deviation)
<u>Communication Scale (visit-specific)</u>	
My health professional's ability to explain things clearly	4.33 (0.84)
My health professional's ability to help me feel comfortable talking about my concerns	4.32 (0.86)
My health professional's ability to take what I say seriously	4.34 (0.85)
The chance to ask all of my questions	4.26 (0.91)
My health professional's ability to answer questions in a sensitive and caring way	4.32 (0.85)
My health professional's willingness to explain different options for my care	4.07 (0.95)
My health professional's interest in how my life affects my health	3.93 (1.02)
The amount of time I had to talk with my health professional	4.12 (0.92)
<u>Administration and Office Procedures Scale (visit-specific)</u>	
The staff's flexibility in scheduling my appointment around my needs	4.01 (0.95)
The courtesy of the office staff	4.06 (0.89)
Help with scheduling my next visit	3.92 (0.93)
How well the staff kept me informed about the waiting time	3.31 (1.21)
Privacy when talking to the receptionist	3.60 (1.00)
The chance to talk to my health professional with my clothes on	4.12 (0.94)
<u>Care Coordination and Comprehensiveness (during past 12 months)</u>	
How well my health care fits my stage of life	3.81 (0.98)
The health professional's interest in my mental and emotional health	3.86 (1.01)
The information I get about healthy living (such as diet and exercise)	3.70 (1.03)
How well the health professionals explain the results of tests or procedures	3.95 (1.00)
The health professional's focus on prevention	3.84 (0.94)
The health professional's knowledge of women's health issues	4.02 (0.91)
Help with finding information resources in women's health	3.62 (1.06)
The chance to get both gynecological and general health care here	3.87 (1.00)
Information about how to get the results of my tests	3.81 (1.04)
My overall trust in the health professionals here	4.18 (0.91)

* Items were scored on a 1-5 scale, with a higher score indicating higher satisfaction.

Table 6. Descriptive Statistics for the Satisfaction Measures (n = 1,202)

	Mean (Standard deviation)	Range of scores	Percent of respondents with highest score	Coefficient alpha
<u>PCSSW Scales</u>				
Communication	33.7 (6.42)	8 – 40	26.1	.96
Administration and Office Procedures	23.0 (4.69)	6 – 30	13.4	.88
Care Coordination and Comprehensiveness	38.6 (8.22)	10 – 50	12.8	.95
<u>Generic Scales</u>				
MOS Visit Satisfaction*	35.9 (6.61)	13 – 45	10.3	.90
Overall Quality of Visit**	4.24 (0.85)	1-5	46.4	NA
<u>CAHPS Quality of Care Rating***</u>	<u>8.60 (1.52)</u>	<u>1 – 10</u>	<u>37.13</u>	<u>NA</u>

PCSSW: Primary Care Satisfaction Survey for Women

MOS: Medical Outcomes Study

CAHPS: Consumer Assessment of Health Plans Study

* 9-item visit-specific satisfaction scale from the Medical Outcomes Study (see text)

** Single item rating of overall quality of care at the visit (see text)

*** Single item rating of overall health care in the last 12 months from the Consumer Assessment of Health Plans Study (see text)

Table 7. Convergent Validity of the PCSSW Scales (unadjusted Pearson correlations; n = 1,202)*

	PCSSW Scales		
	Communication	Administration and Office Procedures	Care Coordination and Comprehensiveness
PCSSW: Communication	---		
PCSSW: Administration and Office Procedures	.6590	---	
PCSSW: Care Coordination and Comprehensiveness	.7923	.6734	---
MOS Visit Satisfaction	.7336	.6710	.7334
Overall Quality of Visit	.7444	.5980	.7047
CAHPS Quality of Care	.5182	.4221	.6081

PCSSW: Primary Care Satisfaction Survey for Women
MOS: Medical Outcomes Study
CAHPS: Consumer Assessment of Health Plans Study

*All correlations are significant ($p < .001$)

Table 8. Discriminant Validity of the PCSSW and Generic Scales (adjusted means and standard errors; n = 1,202)*

	PCSSW			MOS Visit Satisfaction	CAHPS Quality of Care
	Communication	Administration/ Office	Coordination & Comprehensiveness		
Length of time at this place					
2 years or longer (n=629)	33.34(0.43)	23.15(0.32)	38.70(0.55)	35.42(0.44)	8.58(0.10)
Less than 2 years (n=573)	32.79(0.43)	23.17(0.31)	37.55(0.55)	34.81(0.44)	8.42(0.10)
p-value (% variance)	NS (0%)	NS (0%)	.02 (0%)	NS (0%)	NS (0%)
Saw Regular Doctor Today					
Yes (n=721)	33.62(0.40)	23.41(0.30)	38.82(0.52)	35.52(0.42)	8.60(0.09)
No (n=473)	32.03(0.46)	22.67(0.34)	36.81(0.59)	34.35(0.48)	8.35(0.11)
p-value (% variance)	<.0001 (1%)	.01 (0%)	<.0001 (1%)	.0050(1%)	.0070(1%)
Counseling topics					
At Least 1 (n=745)	33.82(0.40)	23.36(0.30)	39.21(0.51)	35.63(0.42)	8.66(0.09)
None (n=457)	31.72(0.45)	22.79(0.33)	36.17(0.57)	34.19(0.46)	8.26(0.10)
p-value (% variance)	<.0001 (2%)	.04 (0%)	<.0001 (3%)	.0003 (1%)	<.0001 (2%)
Preventive screening services					
High (3 or more) (n=343)	33.80(0.48)	23.51(0.35)	39.53(0.62)	35.49(0.49)	8.63(0.11)
Low (n=859)	32.75(0.40)	23.01(0.29)	37.57(0.51)	34.95(0.41)	8.46(0.09)
p-value (% variance)	.01 (0%)	NS (0%)	.0005 (1%)	NS (0%)	.06 (0%)
Got everything needed					
Yes (n=1,138)	33.49(0.37)	23.27(0.28)	38.50(0.48)	35.37(0.39)	8.58(0.09)
No (n=53)	24.26(0.91)	20.50(0.69)	29.37(1.19)	29.19(0.96)	7.17(0.21)
p-value (% variance)	<.0001 (9%)	<.0001(1%)	<.0001 (5%)	<.0001(3%)	<.0001 (4%)

PCSSW= Primary Care Satisfaction Survey for Women

MOS=Medical Outcomes Study

CAHPS= Consumer Assessment of Health Plans

* Means are adjusted for site, age, education, and perceived health status.

** “At your visit today, did you get everything that you thought you needed?”

*** % variance is the percent of the variance in the satisfaction scale that is explained by group membership. This % is also referred to as the eta-squared (see methods).

Table 9. Predictive Validity: Relationship of the Satisfaction Measures to Behavioral Intentions and Self-efficacy for Preventive Care (adjusted means and standard errors; n = 1,202)*

	Communication	PCSSW Administration / Office	Coordination & Comprehensiveness	MOS Visit Satisfaction	CAHPS Quality of Care
Plan to return to office					
Definitely Yes (n=1,096)	33.73(0.37) ¹	23.55(0.28)	39.10(0.47)	35.91(0.37)	8.68(0.08)
Other (n= 105)	26.72(0.68)	19.50(0.51)	28.80(0.85)	27.57(0.68)	6.88(0.15)
p-value (% variance)	<.0001(10%)	<.0001(6%)	<.0001(12%)	<.0001(12%)	<.0001(11%)
Recommend this office/clinic					
Definitely Yes (n=958)	34.49(0.35)	23.97(0.27)	40.09(0.44)	36.64(0.36)	8..83(0.08)
Other (n=240)	27.56(0.48)	19.90(0.37)	30.35(0.61)	29.98(0.49)	7.25(0.11)
p-value (% variance)	<.0001(19%)	<.0001(12%)	<.0001(22%)	<.0001(21%)	<.0001(18%)
Plan to follow advice					
Definitely Yes (n=973)	33.92(0.37)	23.53(0.28)	39.11(0.49)	35.91(0.39)	8.64(0.09)
Other (n=228)	29.40(0.56)	21.40(0.42)	34.16(0.73)	31.72(0.58)	7.92(0.13)
p-value (% variance)	<.0001(6%)	<.0001(3%)	<.0001(5%)	<.0001(5%)	<.0001(3%)
Want to see same professional					
Definitely Yes (n=983)	34.24(0.36)	23.64(0.28)	39.43(0.47)	36.12(0.38)	8.73(0.08)
Other (n=215)	27.48(0.51)	20.87(0.40)	31.97(0.67)	30.32(0.54)	7.51(0.12)
p-value (% variance)	<.0001(17%)	<.0001(5%)	<.0001(12%)	<.0001(11%)	<.0001(9%)
Self efficacy for preventive care					
High (Score of 20+) (n=606)	34.79(0.44)	24.06(0.33)	40.56(0.56)	37.08(0.45)	8.84(0.10)
Low (n=588)	31.94(0.40)	22.57(0.30)	36.52(0.52)	33.83(0.41)	8.30(0.09)
p-value (% variance)	<.0001(4%)	<.0001(2%)	<.0001(5%)	<.0001(5%)	<.0001(3%)

PCSSW= Primary Care Satisfaction Survey for Women
MOS=Medical Outcomes Study
CAHPS= Consumer Assessment of Health Plans

* Means are adjusted for site, age, education, and perceived health status.

Table 10. Contribution of the PCSSW Scales and MOS Visit Satisfaction Scale to Overall Ratings of Quality of Care (n = 1,202)

	Overall Quality of Care At This Visit		CAHPS Overall Quality Rating	
	R ²	Model p-value	R ²	Model p-value
Patient covariates ^a	1%	.0483	8%	<.0001
MOS Visit Satisfaction Scale ^b	50%	<.0001	35%	<.0001
PCSSW Scales ^c				
Communication	75%	<.0001	32%	<.0001
Administration/Office Procedures	42%	<.0001	25%	<.0001
Care Coordination and Comprehensiveness	59%	<.0001	42%	<.0001

PCSSW: Primary Care Satisfaction Survey for Women

MOS: Medical Outcomes Study

CAHPS: Consumer Assessment of Health Plans Study

^a Patient covariates include site, age, education, and perceived health status.

^b Results are based on linear regression analyses in which overall ratings of quality are regressed on the MOS scale. The p-values pertain to the contribution to variance explained in the outcome by the MOS scale and patient covariates.

^c Results are based on linear regression analyses in which overall ratings of quality are regressed on each PCSSW scale separately. The p-values pertain to the contribution to variance explained in the outcome by the indicated PCSSW scale and patient covariates.

Table 11. Factor Loadings Across Age and Race/ethnic Subgroups

	Parameter	Race/ethnicity			Age		
		White	Black	Other	18-35	36-54	55+
		N=799	N=282	N=120	N=518	N=415	N=264
Items Loading on Communication Scale:							
Q11g: Amount of time to talk with health professional	B1	0.82	0.84	0.78	0.84	0.82	0.80
Q11h: Health professional's ability to answer questions in a sensitive and caring way	B2	0.93	0.94	0.97	0.93	0.95	0.92
Q11i: Health professional's ability to explain things clearly	B3	0.93	0.95	0.96	0.94	0.95	0.93
Q11j: Health professional's ability to help me feel comfortable talking about my concerns	B4	0.95	0.92	0.97	0.94	0.95	0.94
Q11k: Chance to ask all of my questions	B5	0.91	0.87	0.92	0.90	0.92	0.89
Q11l: Health professional's ability to take what I say seriously	B6	0.92	0.91	0.93	0.91	0.92	0.95
Q11o: Health professional's willingness to explain different options for my care	B7	0.81	0.80	0.84	0.80	0.80	0.87
Q11p: Health professional's interest in how my life affects my health	B8	0.77	0.74	0.81	0.73	0.77	0.84
Items Loading on Administration and Office Procedures Scale:							
Q11a: Courtesy of the office staff	B9	0.80	0.85	0.83	0.83	0.83	0.77
Q11b: Staff's flexibility in scheduling my appointment around my needs	B10	0.82	0.80	0.84	0.80	0.81	0.85
Q11c: Privacy when talking to the receptionist	B11	0.72	0.70	0.79	0.76	0.69	0.70
Q11d: How well staff kept me informed about the waiting time	B12	0.69	0.67	0.59	0.67	0.68	0.68
Q11e: Help with scheduling next visit	B13	0.79	0.85	0.86	0.81	0.80	0.84
Q11f: Chance to talk to my health professional with my clothes on	B14	0.62	0.75	0.56	0.64	0.60	0.75
Items Loading on Care Coordination and Comprehensiveness Scale:							
Q12a: Health professional's focus on prevention	B15	0.85	0.79	0.83	0.80	0.85	0.86
Q12b: Health professional's knowledge of women's health issues	B16	0.84	0.84	0.83	0.82	0.86	0.87
Q12c: Information I get about healthy living	B17	0.84	0.77	0.85	0.79	0.85	0.84
Q12d: Health professional's interest in my mental and emotional health	B18	0.86	0.83	0.89	0.83	0.88	0.86
Q12e: Help finding information resources in women's health	B19	0.81	0.81	0.82	0.82	0.80	0.82
Q12f: How well my health care fits my stage of life	B20	0.88	0.85	0.91	0.88	0.86	0.89
Q12i: How well health professionals explain results of tests or procedures	B21	0.81	0.83	0.86	0.82	0.79	0.85
Q12h: Information about how to get the results of my tests	B22	0.76	0.83	0.73	0.79	0.74	0.79
Q12l: The chance to get both gynecological and general health care here	B23	0.64	0.78	0.56	0.67	0.61	0.74
Q12m: Overall trust in health professionals here	B24	0.81	0.80	0.78	0.79	0.80	0.84

Table 12. Descriptive Statistics and Pearson Correlations of Path Analysis Variables
(n = 1, 021)

Measure	Mean	SD	1	2	3	4	5	6	7	8	9	10
1. Perceived Health Status	3.4	1.0										
2. Expectations Discrepancy	1.0	0.6	.02									
3. Site Continuity	0.7	0.3	.10	.00								
4. Provider Continuity	1.2	0.7	.06	.01	-.09							
5. PCSSW Communication Scale	33.8	6.4	.11	.69	.01	.05						
6. PCSSW Office Administration and Procedures Scale	23.0	4.6	.04	.64	.05	-.01	.66					
7. PCSSW Care Coordination and Comprehensiveness Scale	38.8	8.2	.09	.63	.05	.01	.79	.67				
8. CAHPS Rating (past 12 months)	8.6	1.5	.04	.41	.05	.01	.52	.43	.61			
9. Visit Rating	4.2	0.8	.09	.58	.02	.01	.75	.60	.71	.60		
10 Age	42.76	16.8	-.30	.05	-.13	-.01	.06	.00	.06	.20	.13	

SD = standard deviation

Table 13. Indices of Fit for the Path Analysis Models (n = 1,021)

	Df	AIC	ECVI	BMD	RMSEA	Cmin/df	χ^2	NFI	CFI
Independence Model	36	27033	26.5	-	0.86	750	27017*	0.0	0.0
Model 1 (uncorrelated errors)	17	936.6	0.92	918	0.22	52	883*	0.97	0.97
Model 2 [BEST] (correlated errors)	13	77.8	0.08	58	0.02	1.2	15.8	1.0	1.0
Saturated Model	0	88.0	0.09	-	-	-	0.0	1.0	1.0

* p < .001

Df: Model degrees of freedom

AIC: Aikake Information Criterion

ECVI: Expected Cross-Validation Index

BMD: Bootstrap Mean Difference, estimates difference between population and model moments.

RMSEA: Root Mean Square Error of Approximation

Cmin/df: Minimum discrepancy divided by model degrees of freedom

χ^2 : Model chi-square from discrepancy function

NFI: Normed Fit Index

CFI: Comparative Fit Index

Table 14. Standardized Regression Weights, Correlations, and Squared Multiple Correlations for the Best Path Model (n = 1,021)

From	To	Standardized Regression Estimate	95% C.I. Lower Bound*	95% C.I. Upper Bound*	Squared Multiple Correlations*
Expectations Discrepancy	CCC	0.628	0.580	0.664	0.396
Site Continuity	CCC	0.046	0.009	0.084	
Expectations Discrepancy	Administration/ Office Procedures	0.642	0.596	0.675	0.412
Expectations Discrepancy	Communication	0.697	0.650	0.727	0.488
Perceived Health Status	Communication	0.057	0.025	0.096	
CCC	CAHPS Rating	0.541	0.469	0.642	0.382
Communication	CAHPS Rating	0.094	0.001	0.180	
Administration/ Office Procedures	Visit Rating	0.129	0.065	0.184	0.602
Communication	Visit Rating	0.464	0.366	0.534	
CCC	Visit Rating	0.252	0.175	0.343	

* 95% Confidence Intervals using the biased corrected percentile bootstrap method (200 subsamples used)

CCC: Care Coordination and Comprehensiveness Scale

APPENDIX A

Final Version of the PCSSW

1. We are interested in your opinions about **your visit today** and about the care you received from the health professionals (the doctors and nurses) and staff. Please rate each of the following things about this visit. (Mark one answer for each item.)

	Not at all Satisfied	Somewhat satisfied	Satisfied	Very satisfied	Extremely satisfied
a. The courtesy of the office staff.....	O	O	O	O	O
b. The staff's flexibility in scheduling my appointment around my needs.....	O	O	O	O	O
c. Privacy when talking to the receptionist..	O	O	O	O	O
d. How well the staff kept me informed about the waiting time.....	O	O	O	O	O
e. Help with scheduling my next visit.....	O	O	O	O	O
f. The chance to talk to my health professional with my clothes on.....	O	O	O	O	O
g. The amount of time I had to talk with my health professional.....	O	O	O	O	O
h. My health professional's ability to answer questions in a sensitive and caring way.....	O	O	O	O	O
i. My health professional's ability to explain things clearly.....	O	O	O	O	O
j. My health professional's ability to help me feel comfortable talking about my concerns....	O	O	O	O	O
k. The chance to ask all of my questions.....	O	O	O	O	O
l. My health professional's ability to take what I					

- say seriously.....
- m. My health professional's willingness to explain different options for my care.....
- n. My health professional's interest in how my life affects my health.....

2. Now please rate all of the care you have received at this office or clinic **during the last 12 months.** If this is your first visit here, please tell us about today. (Mark one answer for each item.)

- | | <u>Not at all</u>
<u>satisfied</u> | <u>Somewhat</u>
<u>satisfied</u> | <u>Satisfied</u> | <u>Very</u>
<u>satisfied</u> | <u>Extremely</u>
<u>satisfied</u> |
|--|---------------------------------------|-------------------------------------|-----------------------|---------------------------------|--------------------------------------|
| a. The health professionals' focus on prevention | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| b. The health professionals' knowledge of women's health issues..... | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| c. The information I get about healthy living (such as diet and exercise)..... | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| d. The health professionals' interest in my mental and emotional health..... | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| e. Help with finding information resources in women's health..... | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| f. How well my health care fits my stage of life... | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| g. Information about how to get the results of my tests..... | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| h. How well the health professionals explain the results of tests or procedures..... | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| i. The chance to get both gynecological and general health care here..... | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| j. My overall trust in the health professionals here..... | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |