

## THE DEVELOPMENT AND VALIDATION OF THE PRIMARY CARE SATISFACTION SURVEY FOR WOMEN

Sarah Hudson Scholle, DrPH<sup>\*a</sup>, Carol S. Weisman, PhD<sup>b</sup>, Roger T. Anderson, PhD<sup>c</sup>, and Fabian Camacho, MS<sup>c</sup>

<sup>a</sup>*Departments of Psychiatry, Pediatrics and Health Services Administration, University of Pittsburgh, Pittsburgh, Pennsylvania*

<sup>b</sup>*University of Michigan School of Public Health, Ann Arbor, Michigan*

<sup>c</sup>*Department of Public Health Sciences, Wake Forest University School of Medicine, Winston-Salem, North Carolina*

---

Assessing patient satisfaction with health care is becoming an integral component of quality monitoring in health care systems, but existing tools typically were developed to minimize differences related to gender. This paper reports the development and psychometric properties of a new survey instrument to measure women's satisfaction with their primary care. A multisite, cross-sectional validation survey of 1,202 women receiving care in primary care settings in Michigan, North Carolina, and Pennsylvania was conducted. Item response theory (IRT) and factor analysis methods were used to identify three scales in the Primary Care Satisfaction Survey for Women (PCSSW): Communication, Administration and Office Procedures, and Care Coordination and Comprehensiveness. Internal consistency reliability is reported, as well as convergent validity in relation to two generic measures (Medical Outcomes Study [MOS] Visit Satisfaction and Consumer Assessment of Health Plans Survey [CAHPS] overall quality rating); discriminant validity in relation to groups expected to have differing satisfaction levels based on previous literature; predictive validity in relation to behavioral intentions; and explanatory power in overall satisfaction ratings. 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.

*Keywords:* Satisfaction; Primary care; Women's health; Item response theory

---

### Background

Existing studies find conflicting effects of gender on patient satisfaction, with some studies reporting that women are more satisfied than men (Cleary, Zaslavsky, & Cioffi, 2000; Weiss, 1988) and others with the opposite results (Kaplan, Sullivan, & Spetter,

1996). A meta-analysis of 110 studies of satisfaction with in- and outpatient care showed no average differences in satisfaction between men and women (Hall & Dornan, 1990), as did a recent analysis of the CAHPS data (Weisman et al., 2001). Yet, several studies suggest that different aspects of care influence outpatient satisfaction ratings among women versus men (Kolodinsky, 1997; Weisman, Rich, Rogers, Crawford, Grayson, & Henderson, 2000; Weisman et al., 2001). These data suggest existing measurement tools may neglect the context of women's primary care experience. Women make more health care visits than men; the majority of patients seen in many primary care settings are women. For that reason alone, information about the quality of their primary care experiences is important. However, women's primary care utilization patterns are complex owing to the structural fragmentation of reproductive and general health care (Clancy & Massion, 1992). To obtain com-

---

Dr. Scholle is now with the National Committee for Quality Assurance, Washington, DC.

Dr. Weisman is now at Penn State College of Medicine, Hershey, Pennsylvania.

Author's note: Copies of the Primary Care Satisfaction Survey for Women (PCSSW) may be obtained at [www.whqa.org](http://www.whqa.org).

Supported by a research grant from the Agency for Healthcare Research and Quality (R01 HS10237-01A1) and contracts from the DHHS Office on Women's Health (Contract no. 00T00215901D).

\* Correspondence to: Sarah Hudson Scholle, DrPH, National Committee for Quality Assurance, 2000 L Street NW, Suite 500, Washington, DC 20036; phone: 202-955-1726.

E-mail: [scholle@ncqa.org](mailto:scholle@ncqa.org).

prehensive care, many women use two physicians (a generalist and an obstetrician/gynecologist) for their regular care (Henderson, Weisman, & Grason, 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 & 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.

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 Survey (CAHPS), which is 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, Henderson, Schifrin, Romans, & Clancy, 2001).

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" 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 specifically include obstetrician/gynecologists in its definition of a 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 confusing or ambiguous, with unknown effects on their responses.

#### *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 Evaluation Working Group reviewed existing outpatient satisfaction tools and concluded that they may not be sensitive to the types of health care innovations intended by the CoE program (Anderson et al., 2001; Scholle, Weisman, Anderson, Weitz, Freund, & Binko, 2000) and set out to develop a tool to capture the totality of women's

primary care experiences. Importantly, the effort was not to develop a tool to measure gender differences, because such a tool 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. With initial funding from the U.S. Department of Health and Human Services Office on Women's Health, six 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 race/ethnicity and socioeconomic status, recruited from multiple communities. 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 five aspects of women's health care: accessing care, privacy and comfort, communication with providers, comprehensiveness of care, and follow-up care (Scholle et al., 2000). 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 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 ( $n = 16$ ) and items pertaining to overall health care at the site during the past year ( $n = 17$ ).

The purpose of this manuscript is to report on the final development and validity of a new tool for assessing women's satisfaction with primary care, the Primary Care Satisfaction Survey for Women (PCSSW). This manuscript presents the results of a multisite field test involving 1,202 women receiving care in diverse primary care settings. In Part I, we describe the development of the final version PCSSW through factor analysis and item response theory (IRT) methods. In Part II, we present data on the reliability and validity of the PCSSW. Readers who are not interested in the details of scale development may prefer to turn directly to Part II.

## Methods

The study consisted of a survey of 1,202 women making primary care visits at sites affiliated with the three participating health centers. The women completed self-administered questionnaires in the sites before and immediately after a primary care visit.

### Setting

Survey data were collected at primary care sites affiliated with the University of Michigan, the University of Pittsburgh, and Wake Forest University. These sites were selected to produce a diverse sample in terms of geographic region, population density, patient demographics, clinical setting (obstetrics/gynecology, internal medicine, and family medicine) and staffing (residents, primary care and specialist physicians, and advanced practice nurses or physician assistants).

### Sample

Women were eligible to participate in the study if they were age 18 or 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. We defined a *primary care visit* as a visit for a routine checkup, gynecologic 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.

*Data collection procedures.* 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. 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. Before seeing their provider, participants completed a pre-visit survey on demographics, visit history, reasons for present visit, and expectations for this visit. After the visit, participants completed a postvisit survey on services received during the visit, satisfaction, whether care met expectations, and additional demographic items. Reasons for nonparticipation were recorded on a daily log sheet. 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. The reasons for declining participation were not having time to remain after the visit to complete the postvisit questionnaire (55%), lack of interest (24%), being too ill to participate (14%), and no reason given (7%). As shown in Table 1, the study group was diverse in socioeconomic and health status.

## Part I. Development

### Scale development methods

To select the final version of the PCSSW, we used a three-step process employing factor analytic methods and IRT (Horn, 1965; Hu & 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 (e.g., child care needs, alternative therapies) were removed before these analyses were conducted.

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 (1965) parallel analysis criterion, conducting Velicer's (1976) minimum average partial procedure, and noting the number of eigenvalues above the average value.

Second, an item elimination and selection process was performed for each factor to develop indicative scales. Items were assessed by considering the magnitude of their factor loadings, assessing the item-total correlation, and fitting Samejima's (1997) graded IRT model to each set of candidate items (using the multi-log program). 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), which are parametric curves that describe the probability (on a scale of 0.0 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. The ideal for ordinal response sets, like those in the PCSSW, would be for respondents with low satisfaction to have a high

**Table 1.** Characteristics of the pooled study sample ( $n = 1,202$ )

Demographic and Health Characteristics	
Age (mean $\pm$ SD)	41.81 $\pm$ 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 <sup>a</sup> (%)	
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 <sup>b</sup>	9.99
Perceived health status <sup>c</sup> (%)	
Excellent	13.42
Very good	35.00
Good	34.33
Fair	14.50
Poor	2.75
Pregnant in past year (%)	18.58
Health Care Utilization	
Number of office visits, past year (mean $\pm$ SD)	7.79 $\pm$ 8.24
Length of time coming to clinical site (%)	
First time today	10.57
<1 year	20.80
1–2 years	16.31
$\geq$ 2 years	52.33
Clinical site is usual source of care (%)	76.44
Main reason for visit (%)	
Follow-up 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 + OB/GYN	40.20
Generalist physician	25.93
OB/GYN	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

Abbreviations: OB/GYN, obstetrician/gynecologist; SD, standard deviation.

<sup>a</sup>“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.

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

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

probability of selecting low category responses and for respondents with high satisfaction to have a high probability of selecting high category responses. The discrimination index (denoted as  $a$ ) is a measure of how well the item response continuum may differentiate respondents among the levels of the latent construct continuum (satisfaction). Items with low index values have substantial overlap among the curves for each response category, whereas items with high index values have trace lines with little or no overlap.

Location parameters are also derived from the ICCs. The number of location parameters is equal to the number of response categories minus one. These location parameters assess item difficulty, which is defined as the point along the measurement of the latent variable (satisfaction) where ICCs intersect among response categories. By convention, the parameter  $b_1$  denotes the point along the continuum for which there is a 50% probability of selecting the lowest response;  $b_2$  denotes the place where there is a 50% probability of selecting either the lowest or next-to-lowest response; and so on. Thus, items with high negative  $b_1$  parameters are better at discriminating among respondents with low satisfaction.

In addition to the ICC of the individual items, IRT modeling also produces information curves for each construct. The information curves assess the contribution of each item in determining the precision with which the overall construct is measured. Unlike classical test theory, in which precision is associated with item reliability, IRT views precision as a function of the level of the latent construct (satisfaction) itself. The information of the overall scale is equal to the inverse of the error variance of theta (the overall construct). 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. For example, some items improve the discriminatory power of the scale because they are less skewed than other items.

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, sensitive to the full range of the latent variables, and 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.

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. To assess the stability and generalizability of the

**Table 2.** Exploratory factor loadings ( $n = 601$ )<sup>a</sup>

		Loadings factor 1	Loadings factor 2
<b>First item set: Today's visit</b>			
Items loading on factor 1 (Communication)			
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
Q11n	The chance to get everything I need at this visit	0.72	0.20
Q11g	The amount of time I had to talk with my health professional	0.67	0.25
Q11m	My health professional's knowledge of my medical history	0.59	0.17
Items loading on factor 2 (Administration and Office Procedures)			
Q11b	The staff's flexibility in scheduling my appointment around my needs	0.08	0.78
Q11a	The courtesy of the office staff	0.10	0.76
Q11e	Help with scheduling my next visit	0.07	0.75
Q11d	How well the staff kept me informed about the waiting time	-0.03	0.72
Q11c	Privacy when talking to the receptionist	0.06	0.66
Item loading both factors			
Q11f	The chance to talk to my health professional with my clothes on	0.36	0.40
Percent variance explained by both factors			70%
Interfactor correlation			.61
			Loadings Factor 1
<b>Second Item Set: Care during the past 12 months</b>			
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 gynecologic and general health care here		0.81
Q12h	Information about how to get the results of my tests		0.80
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
Q12m	My overall trust in the health professionals here		0.69
Q12j	The chance to see the same health professional at each visit		0.64
Percent variance explained by factor			64%

<sup>a</sup>Results of a principal factor analysis with a Promax rotation.

proposed factor structure, the same analysis was conducted on the test sample, and results were compared. The  $\chi^2$  test and the Tucker Reliability Index (Hu & Bentler, 1995) were also assessed to evaluate the fit of the new factor model on the test set.

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  $\geq 55$ ), we used multiple group structural equation modeling. 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 compared to the fit indices of the unconstrained model. Good model fit and performance when compared to the unconstrained model indicates that there

is not sufficient evidence to reject the hypothesis of factor invariance (equal regression weights across groups).

### Results

*Items relating to health care at this visit.* Table 2 shows the results of the exploratory factor analysis. In the analysis of the visit-specific items, two factors were retained: Communication and Administration and Office Procedures. 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 perform similarly with

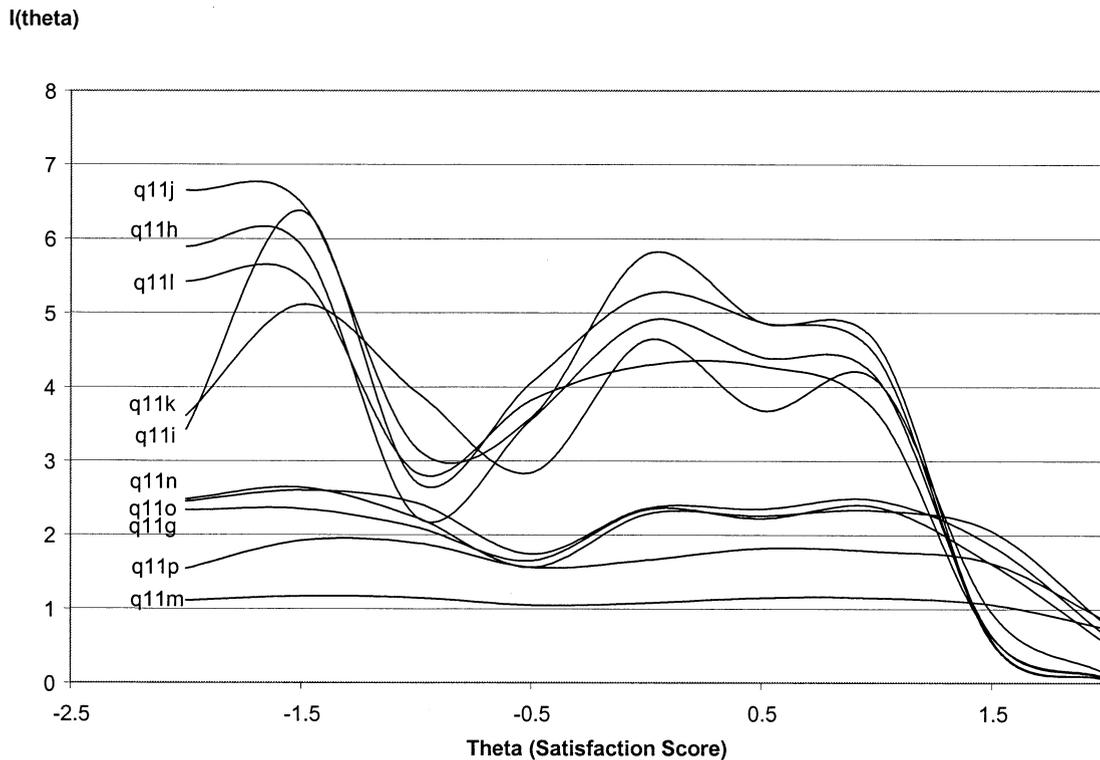
**Table 3.** Item response theory results: Samejima graded model item parameters and standard errors

Abbreviated item content	<i>a</i>	<i>b</i> <sub>1</sub>	<i>b</i> <sub>2</sub>	<i>b</i> <sub>3</sub>	<i>b</i> <sub>4</sub>
<b>First item set: Today's visit</b>					
Factor 1: Communication					
Q11g The amount of time I had to talk with my health professional	3.01 (0.14)	-2.25 (0.17)	-1.27 (0.13)	0.09 (0.05)	1.10 (0.04)
Q11h My health professional's ability to answer questions in a sensitive and caring way	4.73 (0.27)	-2.10 (0.28)	-1.49 (0.08)	-0.19 (0.04)	0.76 (0.03)
Q11i My health professional's ability to explain things clearly	5.09 (0.28)	-2.42 (0.18)	-1.43 (0.06)	-0.21 (0.05)	0.75 (0.03)
Q11j My health professional's ability to help me feel comfortable talking about my concerns	5.13 (0.28)	-2.08 (0.15)	-1.38 (0.08)	-0.16 (0.04)	0.76 (0.03)
Q11k The chance to ask all my questions	4.29 (0.19)	-1.76 (0.10)	-1.22 (0.06)	-0.09 (0.05)	0.83 (0.04)
Q11l My health professional's ability to 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)
Q11m My health professional's 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 at this visit	3.07 (0.16)	-1.95 (0.13)	-1.23 (0.09)	0.18 (0.05)	1.25 (0.04)
Q11o My health professional's willingness to explain different options for my care	3.08 (0.15)	-1.92 (0.16)	-1.14 (0.09)	0.17 (0.04)	1.17 (0.04)
Q11p My health professional's 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)
Factor 2: Administration and Office Procedures					
Q11a The 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 The staff's flexibility in scheduling my appointment around my needs	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 How well the staff kept me 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 my next visit	3.18 (0.18)	-2.02 (0.14)	-1.66 (0.10)	-0.09 (0.04)	0.78 (0.05)
Q11f The chance to talk to my health professional with my clothes on	1.79 (0.12)	-2.60 (0.24)	-2.02 (0.16)	-0.55 (0.07)	0.56 (0.06)
<b>Second item set: Care during past year</b>					
Q12a The health professional's focus on prevention	2.83 (0.13)	-2.20 (0.15)	-1.38 (0.08)	-0.03 (0.04)	1.08 (0.05)
Q12b The health professional's 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 The information I get about healthy living (such as diet and exercise)	2.78 (0.13)	-1.88 (0.10)	-1.09 (0.06)	0.18 (0.04)	1.18 (0.05)
Q12d The health professionals' 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 Help with finding 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 How well my health 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 How well my health 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 Information about how to 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 How well the health professionals explain the 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 The chance to see the 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 The chance to see a 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 The chance to get both gynecologic and general health care here	1.84 (0.10)	-2.46 (0.20)	-1.58 (0.11)	-0.07 (0.06)	1.03 (0.07)
Q12m My overall trust in the health professionals here	2.82 (0.15)	-2.48 (0.20)	-1.59 (0.09)	-0.46 (0.05)	0.52 (0.04)

higher discriminating properties than other items. 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.

Figure 1 shows the information curves for the Communication Scale items, with a cluster of similar performing items (Q11i, Q11j, Q11k, Q11l, and Q11h) distinguishable from less reliable items. This model strongly suggests that these items should be retained.

Of the 10 items that loaded on the Communication factor, 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 his-



**Figure 1.** Information curves for Communication items. The figure illustrates the contribution of each item to the precision with which overall construct of communication is measured. Items Q11i, Q11j, Q11k, Q11l, and Q11h form a cluster that provide more information (i.e., greater precision) to the measurement of communication to the other items.

tory"). 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 five 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 because office procedures generally dictate whether patients are fully clothed when they see their clinicians. Table 3 and Figure 2 describe the discriminatory ability and information curves for this scale. 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.

*Items relating to health care during past year.* For items pertaining to health care during the past year, the factor analysis suggested a single factor solution. 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 3.) 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.

*Confirmatory factor analyses and structural equation modeling.* The confirmatory factor analyses on the exploratory sample explained 70% of the total variance in the visit-specific items and 66% of the

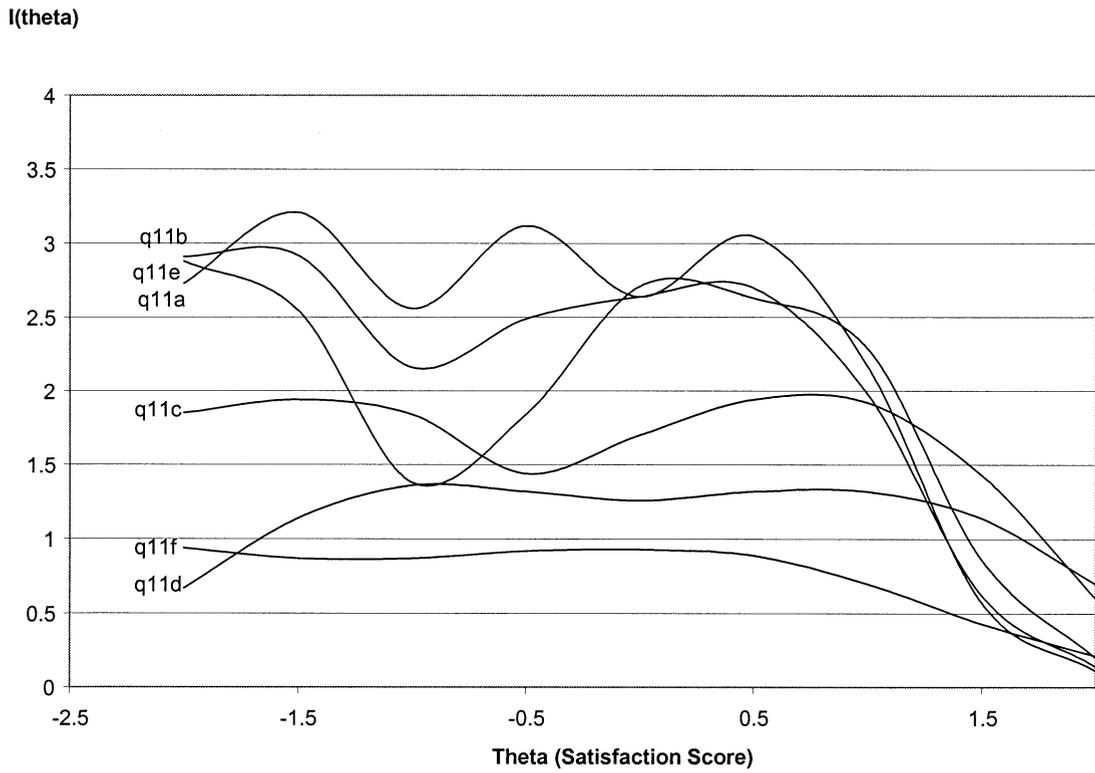


Figure 2. Information curves for Administration and Office procedure items.

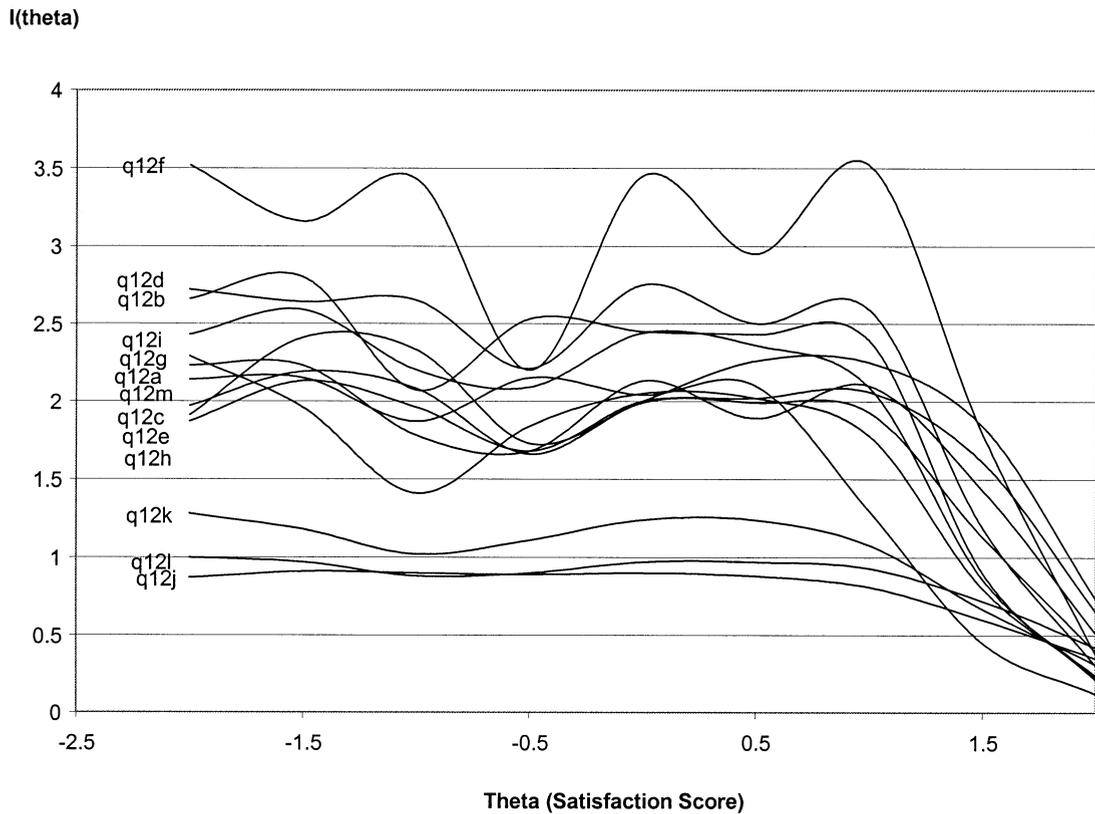


Figure 3. Information curves for Care Coordination and Comprehensiveness items.

**Table 4.** Factor analysis using maximum likelihood factor analysis on both the initial and test data sets

		Loadings factor 1		Loadings factor 2	
		Initial (N = 601)	Test (N = 601)	Initial (N = 601)	Test (N = 601)
<b>First item set: today's visit</b>					
Factor 1: Communication					
Q11j	My health professional's ability to help me feel comfortable talking about my concerns	0.97	0.97	-0.02	-0.03
Q11i	My health professional's ability to explain things clearly	0.98	0.92	-0.04	0.02
Q11l	My health professional's ability to take what I say seriously	0.94	0.90	0.01	0.02
Q11h	My health professional's ability to answer questions in a sensitive and caring way	0.93	0.90	0.03	0.05
Q11k	The chance to ask all my questions	0.92	0.89	0.03	-0.02
Q11o	My health professional's willingness to explain different options for my care	0.70	0.76	0.17	0.11
Q11p	My health professional's interest in how my life affects my health	0.68	0.68	0.11	0.14
Q11g	The amount of time I had to talk with my health professional	0.67	0.72	0.26	0.17
Factor 2: Administration and Office Procedures					
Q11b	The staff's flexibility in scheduling my appointment around my needs	0.08	0.08	0.80	0.72
Q11a	The courtesy of the office staff	0.08	0.20	0.79	0.66
Q11e	Help with scheduling my next visit	0.05	0.09	0.78	0.76
Q11d	How well the staff kept me informed about the waiting time	-0.02	-0.02	0.69	0.72
Q11c	Privacy when talking to the receptionist	0.06	-0.04	0.64	0.83
Q11f	The chance to talk to my health professional with my clothes on	0.33	0.36	0.42	0.40
Percent variance explained by both factors			Initial = 71%; Test = 70%		
Interfactor correlation			Initial = .60; Test = .60		
		Loadings factor 1			
		Initial (N = 601)	Test (N = 601)		
<b>Second item set: care during past year</b>					
Q12f	How well my health care fits my stage of life			0.88	0.89
Q12d	The health professionals' interest in my mental and emotional health			0.88	0.85
Q12c	The information I get about healthy living (such as diet and exercise)			0.85	0.82
Q12b	The health professional's knowledge of women's health issues			0.83	0.85
Q12e	Help with finding information resources in women's health			0.83	0.83
Q12e	Help with finding information resources in women's health			0.83	0.83
Q12i	How well the health professionals explain the results of tests or procedures			0.81	0.82
Q12m	My overall trust in the health professionals here			0.78	0.79
Q12h	Information about how to get the results of my tests			0.77	0.79
Q12l	The chance to get both gynecologic and general health care here			0.65	0.66
Percent variance explained by factor			Initial = 66%; Test = 66%		

variance for the past-year items (Table 4). 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 (also shown in Table 4) 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. The structural equation modeling 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. Similar analyses suggested factor invariance across age groups as well.

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 5-point scale: 1 = not at all satisfied; 2 = somewhat satisfied; 3 = satisfied; 4

**Table 5.** Primary Care Satisfaction Survey for Women: Scales and item univariate statistics<sup>a</sup> (*n* = 1,202)

Scale and item	Mean (SD)
<b>Communication Scale (visit specific)</b>	
Q11g. The amount of time I had to talk with my health professional	4.12 (0.92)
Q11h. My health professional's ability to answer questions in a sensitive and caring way	4.32 (0.85)
Q11i. My health professional's ability to explain things clearly	4.33 (0.84)
Q11j. My health professional's ability to help me feel comfortable talking about my concerns	4.32 (0.86)
Q11k. The chance to ask all of my questions	4.26 (0.91)
Q11l. My health professional's ability to take what I say seriously	4.34 (0.85)
Q11o. My health professional's willingness to explain different options for my care	4.07 (0.95)
Q11p. My health professional's interest in how my life affects my health	3.93 (1.02)
<b>Administration and Office Procedures Scale (visit specific)</b>	
Q11a. The courtesy of the office staff	4.06 (0.89)
Q11b. The staff's flexibility in scheduling my appointment around my needs	4.01 (0.95)
Q11c. Privacy when talking to the receptionist	3.60 (1.00)
Q11d. How well the staff kept me informed about the waiting time	3.31 (1.21)
Q11e. Help with scheduling my next visit	3.92 (0.93)
Q11f. The chance to talk to my health professional with my clothes on	4.12 (0.94)
<b>Care Coordination and Comprehensiveness (during past 12 months)</b>	
Q12a. The health professional's focus on prevention	3.84 (0.94)
Q12b. The health professional's knowledge of women's health issues	4.02 (0.91)
Q12c. The information I get about healthy living (such as diet and exercise)	3.70 (1.03)
Q12d. The health professional's interest in my mental and emotional health	3.86 (1.01)
Q12e. Help with finding information resources in women's health	3.62 (1.06)
Q12f. How well my health care fits my stage of life	3.81 (0.98)
Q12h. Information about how to get the results of my tests	3.81 (1.04)
Q12i. How well the health professionals explain the results of tests or procedures	3.95 (1.00)
Q12l. The chance to get both gynecologic and general health care here	3.87 (1.00)
Q12m. My overall trust in the health professionals here	4.18 (0.91)

<sup>a</sup>Items were scored on a 1–5 scale, with a higher score indicating higher satisfaction.

= 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 nonmissing items imputed if there were fewer than 25% missing items on the scale.

## Part II: Reliability and Validity

### Approach to analysis

Several approaches were used to assess the reliability and validity of the PCSSW scales. For each PCSSW scale and the comparison generic patient satisfaction measures, we present the mean and range. In addition, 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 unadjusted Pearson correlation of the PCSSW scales with the generic tools. 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 ( $\eta^2$ ).  $\eta^2$  is the proportion of the sum of squares attributable to the criterion

variable divided by the total corrected sum of squares (Becker, 1999). An  $\eta^2$  close to 0 implies that the two groups are difficult to distinguish, and an  $\eta^2$  close to 1 implies there is a clear difference between the two groups. Predictive validity was assessed the same way.

To compare the PCSSW scales' ability to capture the variance in quality assessments to the generic Medical Outcomes Study (MOS) scale, 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

**Generic satisfaction tools.** We used three generic measures of satisfaction. The *MOS Patient Visit Satisfaction Scale* is a 9-item scale tapping multiple dimensions of care; items are rated on a 5-point excellent-to-poor

response set (rated 5–1) and summed to give a score. The instrument discriminates between types of practice settings and health plans, and predicts such outcomes as returning for follow-up care and whether patients will change physicians within 6 months (Davies & Ware, 1991; Nelson et al., 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 (HEDIS Protocol, 1999). 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 5-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 & Strasser, 1993; Cleary & McNeil, 1988; Donaldson, 2001). For discriminant validity analyses, two measures of continuity were used: 1) for those using the site as their usual source of care, having longer tenure at this site ( $\geq 2$  years), versus shorter tenure ( $< 2$  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. We also hypothesized that women who saw their regular doctor on the day of the visit would have higher ratings on the Communication scale).

*Comprehensiveness of care.* The comprehensiveness of preventive care is associated with higher satisfaction with care in several studies (Cleary & McNeil, 1988; Orlando & Meredith, 2002; Schauffler, Rodrigues, & Milstein, 1996; Sitzia & Wood, 1997). Three variables reflecting comprehensiveness were defined for the discriminant validity analyses. *Comprehensiveness of preventive screening services* is the number of age-appropriate screening services provided at the visit. For all women, these included blood pressure check, Pap smear, and physical breast examination. For women ages 50 and over, mammogram, blood cholesterol test, and colon cancer screening also were included. The sum of services received was dichotomized as three or more

services provided versus less than three. *Comprehensiveness of preventive counseling* is the number of age-appropriate counseling topics discussed during nonillness 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; and alternative therapies, such as herbal products or massage therapy. Additional topics included preventing unwanted pregnancy or planning a pregnancy (for women ages 18–44) and menopause or hormone replacement (for women ages  $\geq 45$ ). The sum of counseling topics was dichotomized as at least one topic discussed at the visit versus none. 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 would have higher ratings on both visit-based and past-year scales of the PCSSW.

*Behavioral intentions.* Measures of behavioral intent are often used in patient satisfaction studies as proxies for postvisit outcomes when these outcomes cannot be observed directly. Patient satisfaction has been correlated with intention to return to the provider, adhere to providers' recommendations, recommend the provider to others, and similar items (Aharony & Strasser, 1993; Ware & Davies, 1983; Ware & Hays, 1988). To assess predictive validity using behavioral intentions, we asked about their plans to follow health professionals' advice, to return to the office or clinic for care, to recommend the office or clinic to family and friends, as well as their desires to see the same health professional. We compared satisfaction scores of women who rate each item definitely yes versus probably yes, probably not, and definitely not. 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.

## Results

*Reliability and convergent validity.* 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,

**Table 6.** Descriptive statistics for the satisfaction measures ( $n = 1,202$ )

	Mean (SD)	Range of scores	Percent of respondents with highest score	Coefficient alpha
<b>PCSSW scales</b>				
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
<b>Generic scales</b>				
MOS Visit Satisfaction <sup>a</sup>	35.9 (6.61)	13–45	10.3	.90
Overall Quality of Visit <sup>b</sup>	4.24 (0.85)	1–5	46.4	NA
CAHPS Quality of Care Rating <sup>c</sup>	8.60 (1.52)	1–10	37.13	NA

*Abbreviations:* PCSSW, Primary Care Satisfaction Survey for Women; MOS, Medical Outcomes Study; CAHPS, Consumer Assessment of Health Plans Study.

<sup>a</sup>Nine-item visit-specific satisfaction scale from the MOS (see text).

<sup>b</sup>Single-item rating overall quality of care at the visit (see text).

<sup>c</sup>Single item rating overall health care in the last 12 months from the Consumer Assessment of Health Plans Study (see text).

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 (0.67–0.73,  $p < .001$  for all) and with the overall quality of visit rating (0.60–0.74,  $p < .001$  for all) are somewhat higher than correlations with the CAHPS overall rating (0.42–0.61,  $p < .001$ ) (Table 7).

*Discriminant validity.* 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). All of the measures showed significantly higher satisfaction ratings when the regular doctor was seen at the visit, although the magnitude of these differences was small.

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*

*Contribution of the PCSSW to overall quality ratings.* All of the satisfaction measures were significantly associated with women’s behavioral intentions (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. In regression analyses examining the

**Table 7.** Convergent validity of the PCSSW scales (unadjusted Pearson correlations;  $n = 1,202$ )<sup>a</sup>

	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

*Abbreviations:* PCSSW, Primary Care Satisfaction Survey for Women; MOS, Medical Outcomes Study; CAHPS, Consumer Assessment of Health Plans Study.

<sup>a</sup>All correlations are significant ( $p < .001$ ).

**Table 8.** Discriminant validity of the PCSSW and generic scales (adjusted means and standard errors;  $n = 1,202$ )<sup>a</sup>

	PCSSW				
	Communication	Administration/Office	Coordination and Comprehensiveness	MOS Visit Satisfaction	CAHPS Quality of Care
Length of time at this place					
≥2 years ( $n = 629$ )	33.34 (0.43)	23.15 (0.32)	38.70 (0.55)	35.42 (0.44)	8.58 (0.10)
<2 years ( $n = 573$ )	32.79 (0.43)	23.17 (0.31)	37.55 (0.55)	34.81 (0.44)	8.42 (0.10)
<i>p</i> -Value (% variance) <sup>b</sup>	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)
<i>p</i> -Value (% variance)	<.0001 (1%)	.01 (0%)	<.0001 (1%)	.0050(1%)	.0070(1%)
Got everything needed <sup>c</sup>					
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)
<i>p</i> -Value (% variance)	<.0001 (9%)	<.0001(1%)	<.0001 (5%)	<.0001(3%)	<.0001 (4%)
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)
<i>p</i> -Value (% variance)	<.0001 (2%)	.04 (0%)	<.0001 (3%)	.0003 (1%)	<.0001 (2%)
Preventive services					
High (≥3) ( $n = 343$ )	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)
<i>p</i> -Value (% variance)	.01 (0%)	NS (0%)	.0005 (1%)	NS (0%)	.06 (0%)

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

<sup>a</sup>Means are adjusted for site, age, education, and perceived health status.

<sup>b</sup>Percent variance is the percent of the variance in the satisfaction scale that is explained by group membership. This percent is also referred to as the eta-squared (see methods).

<sup>c</sup>“At your visit today, did you get everything that you thought you needed?”

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) contrib-

ute 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

**Table 9.** Predictive validity: Relationship of the satisfaction measures to behavioral intentions and self-efficacy for preventive care (adjusted means and standard errors;  $n = 1,202$ )<sup>a</sup>

	PCSSW				
	Communication	Administration/Office	Coordination and Comprehensiveness	MOS Visit Satisfaction	CAHPS Quality of Care
Plan to return to office					
Definitely yes ( $n = 1,096$ )	33.73 (0.37)	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)
<i>p</i> -Value (% variance) <sup>b</sup>	<.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)
<i>p</i> -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)
<i>p</i> -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)
<i>p</i> -Value (% variance)	<.0001 (17%)	<.0001 (5%)	<.0001 (12%)	<.0001 (11%)	<.0001 (9%)

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

<sup>a</sup>Means are adjusted for site, age, education, and perceived health status.

<sup>b</sup>Percent variance is the percent of the variance in the satisfaction scale that is explained by group membership. This percent is also referred to as the eta-squared (see methods).

**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 <sup>a</sup>	1%	.0483	8%	<.0001
MOS Visit Satisfaction Scale <sup>b</sup>	50%	<.0001	35%	<.0001
PCSSW Scales <sup>c</sup>				
Communication	75%	<.0001	32%	<.0001
Administration/Office Procedures	42%	<.0001	25%	<.0001
Care Coordination Comprehensiveness	59%	<.0001	42%	<.0001

*Abbreviations:* PCSSW, Primary Care Satisfaction Survey for Women; MOS, Medical Outcomes Study; CAHPS, Consumer Assessment of Health Plans Study.

<sup>a</sup>Patient covariates include site, age, education, and perceived health status.

<sup>b</sup>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.

<sup>c</sup>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.

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.

## Discussion

### *Main findings and implications*

The PCSSW is a 24-item survey tool consisting of three scales that have been shown to be psychometrically valid and to have greater explanatory power than an existing generic patient satisfaction scale for women's health care. Developed through careful qualitative work involving women with diverse social and health characteristics across the country, the PCSSW demonstrates factor invariance across subgroups of women defined by age group and by race/ethnicity and strong psychometric properties in the field test.

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 obtain gynecologic 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 PCSSW adds sensitivity to satisfaction measurement and can be useful in evaluations of the quality of primary care and in quality improvement programs, as demonstrated in two recent 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. In adjusted comparisons, women served in VA women's clinics reported significantly higher satisfaction on all five domains (Bean-Mayberry, Chang, McNeil, Whittle, Hayes, & Scholle, 2003). Second, in the national evaluation of the quality of care in the clinical care centers of the National 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, Weisman, Scholle, Henderson, Oldendisk, & Camacho, 2002).

### *Limitations*

This project is 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. The results of this project, which demonstrate the usefulness of the tool, 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 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 evaluative alternative models of primary care delivery for women and in quality improvement programs in women's primary care (Sitzia and Wood, 1997).

## Acknowledgments

We 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, Jillian Henderson, 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, and Rik Lennox, PhD, for his review of the psychometric analyses.

## References

- Aharony, L., & Strasser, S. (1993). Patient satisfaction: What we know about and what we still need to explore. *Medical Care Review*, 50, 49–80.
- Anderson, R. T., Barbara, A. M., Weisman, C., Hudson Scholle, S., Binko, J., Schneider, T., et al (2001). A qualitative analysis of women's satisfaction with primary care from a panel of focus groups in the National Centers for Excellence in Women's Health. *Journal of Womens Health & Gender-Based Medicine*, 10, 637–647.
- Anderson, R. T., Weisman, C. S., Scholle, S. H., Henderson, J. T., Oldendisk, R., & Camacho, F. (2002). Evaluation of the quality of care in the clinical centers of the national centers of excellence in women's health. *Women's Health Issues*, 12, 309–326.
- Bean-Mayberry, B. A., Chang, C. C., McNeil, M. A., Whittle, J., Hayes, P. M., & Scholle, S. H. (2003). Patient satisfaction in women's clinics and traditional primary care clinics in the Veteran's Administration. *Journal of General Internal Medicine*, 18, 175–181.
- Becker, L. A. (1999). *Measures of Effect Size (Strength of Association)*. 1999. Available: [http://web.uccs.edu/lbecker/spss/gml\\_effectsize.htm](http://web.uccs.edu/lbecker/spss/gml_effectsize.htm).
- Clancy, C. M., & Massion, C. T. (1992). American women's health care: A patchwork quilt with gaps. *Journal of the American Medical Association*, 268, 1918–1920.
- Cleary, P. D., & McNeil, B. J. (1988). Patient satisfaction as an indicator of quality care. *Inquiry*, 25, 25–36.
- Cleary, P. D., Zaslavsky, A. M., & Cioffi, M. (2000). Sex differences in assessments of the quality of Medicare managed care. *Women's Health Issues*, 10, 70–79.
- Davies, A. R., & Ware, J. E. Jr (1991). GHAA's Consumer Satisfaction Survey and User's Manual. Washington, DC: Group Health Association of America.
- Donaldson, M. S. (2001). Continuity of care: A reconceptualization. *Medical Care Research and Review*, 58, 255–290.
- Hall, J. A., & Dornan, M. C. (1990). Patient sociodemographic characteristics as predictors of satisfaction with medical care: A meta-analysis. *Social Science and Medicine*, 30, 811–818.
- Hays, R. D., Shaul, J. A., Williams, V. S., Lubalin, J. S., Harris-Kojetin, L. D., Sweeny, S. F., et al (1999). Psychometric properties of the CAHPS 1.0 Survey measures. *Medical Care*, 37, MS22–MS31.
- HEDIS Protocol for Administering CAHPS 20H Survey. (1999). HEDIS. Washington, DC: National Committee for Quality Assurance.
- Henderson, J. T., & Weisman, C. S. (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, J. T., Weisman, C. S., & Grason, H. (2002). Are two doctors better than one? Womens physician use and appropriate care. *Womens Health Issues*, 12, 138–149.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179–185.
- Hu, L.-T., Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.). *Structural Equation Modeling, Concepts, Issues and Applications*. Thousand Oaks, California: Sage.
- Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika*, 35, 401–415.
- Kaplan, S. H., Sullivan, L. M., & Spetter, D.). Gender and patterns of physician-patient communication. In Falik, M. M.; Collins, K. S., editors. *Women's Health: The Commonwealth Fund Survey*. Baltimore, MD: The Johns Hopkins University Press; 1996.
- Kolodinsky, J. (1997). Gender differences in satisfaction with primary care physicians in a managed care health plan. *Women and Health*, 26, 67–86.
- Nelson, E. C., Larson, C. O., Davies, A. R., Gustafson, D., Ferreira, P. L., Ware, J. E., et al (1991). The patient comment card: A system to gather customer feedback. *Quality Review Bulletin*, 17, 278–285.
- Orlando, M., & Meredith, L. S. (2002). Understanding the causal relationship between patient-reported interpersonal and technical quality of care for depression. *Medical Care*, 40, 696–704.
- Rubin, H. R., Gandek, B., Rogers, W. H., Kosinski, M., McHorney, C. A., Ware, J. E., et al (1993). Patients' ratings of outpatient visits in different practice settings. Results from the Medical Outcomes Study. *Journal of the American Medical Association*, 270, 835–840.
- Samejima, F.). Models for items with polytomous response formats: Graded response model. In Linden, W. J.; Hambleton, R. K., editors. *Handbook of Modern Item Response Theory*. New York: Springer-Verlag; 1997.
- Schauffler, H. H., Rodriguez, T., & Milstein, A. (1996). Health education and patient satisfaction. *The Journal of Family Practice*, 42, 62–68.
- Scholle, S. H., Weisman, C. S., Anderson, R., Weitz, T., Freund, K. M., & 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, 1–9.
- Sitzia, J., & Wood, N. (1997). Patient satisfaction: A review of issues and concepts. *Social Science and Medicine*, 45, 1829–1843.
- Velicer, W. F. (1976). The relation between factor score estimates, image scores, and principal component scores. *Educational and Psychological Measurement*, 36, 149–159.
- Ware, J. E., & Davies, A. R. (1983). Behavioral consequences of consumer dissatisfaction with medical care. *Evaluation and Program Planning*, 6, 291–297.

- Ware, J. E., & Hays, R. D. (1988). Methods for measuring patient satisfaction with specific medical encounters. *Medical Care*, 26, 393–402.
- Ware, J., & Sherbourne, C. (1992). The MOS 36-Item Short-Form Health Survey SF-36 I. Conceptual Framework and Item Selection. *Medical Care*, 30, 473–483.
- Weisman, C. S., Rich, D. E., Rogers, J., Crawford, K. G., Grayson, C. E., & Henderson, J. T. (2000). Gender and patient satisfaction with primary care: Tuning in to women in quality measurement. *Journal of Womens Health and Gender-Based Medicine*, 9, 657–665.
- Weisman, C. S., Henderson, J. T., Schifrin, E., Romans, M., & Clancy, C. M. (2001). Gender and patient satisfaction in managed care plans: Analysis of the 1999 HEDIS/CAPHS 2.0H Adult Survey. *Women's Health Issues*, 11, 401–415.
- Weiss, G. L. (1988). Patient satisfaction with primary medical care. *Medical Care*, 26, 383–392.

---

### Author Description

Dr. Hudson Scholle is a health services researcher with an interest in quality of health care for women and families.

Dr. Weisman is a sociologist and health services researcher with a principal interest in women's health care and policy.

Dr. Anderson is a health services researcher who conducts research on patient satisfaction and health care quality.

Mr. Camacho is a research associate with an interest in health-related quality of life and scale development.

---