
Introduction

This thesis has two central themes, racial/ethnic differences in consumer evaluations of care and methodological concerns related to making those comparisons. The first theme, racial/ethnic differences in patient evaluations of care, is addressed in chapters 4 to 6 of this thesis. The second theme, methodological concerns related to making racial/ethnic comparisons of evaluations of care, is addressed in Chapters 2, 3, 7.

The two principal purposes of this chapter are to summarize the results from the main substantive and methodological chapters of this thesis and to present the implications of those results for health policy and future research. With these objectives in mind, the remainder of this chapter is organized into three sections. In the next section, the main results from the substantive chapters are summarized. In the third section of this chapter, the main results from the methodological chapters are summarized. In the fourth section, the implications of the main substantive and methodological results for health policy and future research are presented.

Summary of Substantive Results

Chapter 4: Are Latinos Less Satisfied with Communication by Health Care Providers?

In Chapter 4, patient ratings of communication by medical providers among non-Hispanic whites, Hispanics who completed their surveys in English, and Hispanics who completed their surveys in
Spanish are compared. The survey data was collected from patients receiving care from the United Medical Group Association (UMGA), an association of physician groups located in the western United States. Of the 6,211 surveys analyzed in this study, 713 were completed by Hispanics, of which 181 were completed in Spanish. The overall survey response rate for the study was 59%.

Patient’s evaluation of communication by medical providers was assessed using a 5-item scale. All five items were administered using an identical 7-point response format (Very Poor, Poor, Fair, Good, Very Good, Excellent, and The Best) along with the option, Does Not Apply to Me. The five items asked survey respondents to rate the following aspects of communication: medical staff listening to what you have to say; answers to your questions; explanations about prescribed medications; explanations about medical tests and procedures; and reassurance and support from your doctor and support staff.

Each rating question was modeled separately using ordinal logistic models. The main independent variables in the study were ethnicity (white versus Hispanic) and language of survey response (English versus Spanish). These two variables were combined to produce three mutually exclusive categories of respondents: whites who responded to the English survey version (whites), Hispanics who responded to the English survey version (E-Hispanics), and Hispanics who responded to the Spanish survey version (S-Hispanics).

The case-mix variables included in all models were age and gender. Other case-mix variables including health status, education, and a Spanish language response variable were included in preliminary models. However, because the results from models using these
additional case-mix variables did not differ from models that only included age and gender, the more parsimonious models were selected for presentation. Huber and White corrected standard errors were computed to correct for intra physician group clustering of survey respondents.

The five final models showed similar patterns of results. Across all five models, S-Hispanics rated communication with their medical providers the lowest, followed by E-Hispanics and whites. Specifically, there were statistically significant differences in ratings between whites and S-Hispanics across all models at the 0.05 level. There were also statistically significant differences between S-Hispanics and E-Hispanics across all questions at the 0.05 level. In addition, there were statistically significant differences between whites and E-Hispanics on three of five questions.

These results suggest that Hispanics who primarily communicate in Spanish are at increased risk for poor communication with their medical providers and subsequent sub-optimal outcomes of care. For example, Spanish speaking patients receiving unsatisfactory explanations about how to take their prescribed medications may inadvertently take them inappropriately, resulting in less than optimal outcomes including medication toxicities, regardless of whether or not the prescriptions were technically appropriate.

The results from this study are consistent with prior research on quality of care for Hispanic patients. Baker et al. (1996) found poor communication between Spanish speaking patients and their medical providers in emergency room settings (Baker, Parker, Williams, Coates, & Pitkin, 1996). Perez-Stable et al. (1999) found worse outcomes of care among Spanish speaking Hispanic patients with non-Spanish speaking
providers than English speaking Hispanic patients (Perez-Stable, Naapoles-Springer, & Miramontes, 1997). The results presented in Chapter 4 are consistent with language barriers faced by all patients with poor English language skills accessing the medical care system (Woloshin, Bickell, Schwartz, Gany, & Welch, 1995b).

Chapter 5: Differences in CAHPS® Adult Survey Ratings and Reports by Race and Ethnicity: An Analysis of the National CAHPS® Benchmarking Data 1.0.

In Chapter 5, a study of adult racial/ethnic differences in consumer evaluations of care using the National CAHPS® Benchmarking Database 1.0 (NCBD 1.0) is presented. The NCBD 1.0 is an aggregation of CAHPS® 1.0 survey results collected by health plans located across the United States. The NCBD project is administered by Quality Measurement Advisory Service (QMAS) with finding from the Agency for Healthcare Quality and Research (AHRQ). Health plans administering the CAHPS® surveys were asked to voluntarily contribute their survey results to the NCBD project for purposes of benchmarking and research. Both adult and child survey results from Medicaid and commercial settings are included in NCBD 1.0.

Although different versions of the CAHPS® surveys have been developed for adults and children, Medicaid and commercial settings, and mail and telephone administration, all versions of the surveys contain the a core set of evaluations. These evaluations can be divided into reports and ratings. Reports ask survey respondents about the frequency with which certain events take place. In the CAHPS® paradigm, there are five reporting domains, each addressed by several
questions. The five reporting domains are access to needed care (four questions), provider communication (four questions), office staff helpfulness (three questions), promptness of needed care (four questions), and health plan customer service (two questions). With the exception of two questions which are asked using a Yes/No format, all questions in the reports domains are asked using a Never, Sometimes, Usually, Always format.

Ratings, in contradistinction to reports, are single items that ask consumers to make summary evaluations of various aspects of care. The CAHPS® 1.0 surveys include four rating items, which ask consumer to rate their personal medical provider, specialists, health care, and health plan.

All four rating questions are administered using an 11-point numeric scale ranging from 0 to 10. Statements like "The Best Possible Personal Doctor" and "The Worst Possible Personal Doctor" anchor the extremes of each scale.

Each report and rating was modeled separately using OLS regression. The main independent variables were indicators of race/ethnicity. Based on two questions on ethnicity (Hispanic or non-Hispanics) and race (White, Black/African American, Asian/Pacific Islander, American Indian/Native Alaskan, Other Race, Multiple Races), seven mutually exclusive respondent categories were created: Hispanics regardless of race and non-Hispanic White, Black/African American, Asian/Pacific Islander, American Indian/Native Alaskan, other or multiple race, and missing race data. Other independent variables included in all models were age, gender, education, and health status.
Because of the skewness of the distributions of the dependent variables, a linear transformation was used to achieve greater normality. However, the regression results varied little whether or not the transformed dependent variables were used, thus only results based on the untransformed variables were presented. Huber and White corrected standard errors were computed to adjust for intra-plan clustering. Weights were derived to adjust for differences in response rates across health plans.

Overall, whites reported better care than other racial/ethnic groups. Compared to whites, Hispanics reported worse access to care, promptness of care, and health plan customer service. Asian/Pacific Islanders reported worse care than whites across all five reporting domains. Persons in the Other/Multiracial category reported worse access to care, promptness of care, provider communication, and health plan customer than whites. Persons in the Missing category also reported worse care than whites across most reports domains. American Indians/Alaskan natives were the only group to report care similar to whites.

African Americans, unexpectedly, reported better care than whites in two domains: provider communication and office staff helpfulness. Interestingly, the office staff reporting composite includes a question about the frequency with which the customer was treated with courtesy and respect. If African Americans felt they had experienced more racial discrimination by office staff than whites, it is likely that this question would have captured that experience.

Ratings revealed fewer differences between whites and other race/ethnic groups than reports, and in some cases, ratings
contradicted reports. For example, Hispanics rated their personal
doctor, specialists and their health care similar to whites, and rated
their health plans higher than whites despite reporting worse access to
care and promptness of care than whites. Asian/Pacific Islanders gave
ratings similar to whites across all five rating questions despite
reporting worse care across all report domains. American
Indian/Alaskan Natives rated their personal doctors and specialists
lower than whites despite similar reports about care. Persons in the
Multiracial/Other and Missing categories rated their care lower than
whites, consistent with worse reports about care given by both groups
compared with whites. African Americans rated their health plans
higher than whites, consistent with better reports about care than
whites.

Chapter 6: Racial and Ethnic Differences in Parents’ Assessments
of Pediatric Care in Medicaid Managed Care.

In Chapter 6, a study of consumer evaluations of pediatric care
based on the NCBD 1.0 is presented. Although this study is based on
the same database as the previous study, several important differences
in the data resulted in the decision to conduct separate studies.
First, pediatric CAHPS® 1.0 data is collected by proxy; the adult
parent or guardian of a child respondent is asked to complete the
survey. Second, only the child versions of the CAHPS® 1.0 surveys
collect information about the main language spoken at home. This
variable enabled further analysis of the child data not possible with
the adult data. Third, it was the judgment of the investigators that
the amount of information generated by simultaneous presentation of the
adult and child results would be overwhelming.
As in the analysis of the adult NCBD 1.0 data, the core reports and ratings were chosen as the main dependent variables in this study. Also as in the adult study, seven racial/ethnic categories served as the main independent variables in this study. The categories included: Hispanics regardless of race, and non-Hispanic Whites, Black/African Americans, Asian/Pacific Islanders, American Indian/Native Alaskans, other or multiple races, missing race data. The adult proxy’s race/ethnicity was used rather than the child’s race/ethnicity because the adult proxy’s completed the surveys.

In addition, Hispanics and Asian/Pacific Islanders were subcategorized by language spoken at home. Specifically, Hispanics were subcategorized as speaking Spanish (Hispanic-S) or English (Hispanic-E) at home and Asian/Pacific Islanders were categorized as speaking English (Asian-E) or another language (Asian-O) at home. A small group of Hispanics, who did not indicate what language they spoke at home (N=26), was dropped from the study. Asian/Pacific Islanders who did not indicate what language they spoke at home formed a third Asian/Pacific Islander group (Asian-M).

The case-mix variables used in this analysis were similar to those used in the study of adult evaluations. They included the proxy’s age, the proxy’s gender, the proxy’s education, and the child’s health status.

Separate OLS models were used for each dependent variable. Because results based on transformed and untransformed dependent variables were similar, only results based on the untransformed dependent variables are presented. Huber and White corrected standard errors were computed to correct for intra-plan clustering and
analytical weights were used to account for variations in response rates across plans.

The first main result of this study is that whites reported better experiences with care and higher ratings of care than members of other race/ethnic groups. Hispanics-S and Asian-O reported worse access to care, promptness of care, provider communication, staff helpfulness, and health plan customer service than whites. Asian-M reported worse promptness of care, provider communication, and staff helpfulness. African Americans reported worse access to care, promptness of care, and health plan customer service than whites. American Indians reported worse access to care, promptness of care, provider communication, and health plan customer service than whites. Finally, persons in the Missing and Multiracial/Other categories reported worse access to care, promptness of care, provider communication, and health plan customer service than whites.

Ratings of care closely mirrored reports of care, with whites rating their experiences with providers and services more highly than other racial/ethnic groups. As in the study of the adult data, there were fewer differences between whites and the other groups based on ratings than reports. However, unlike the adult study, there were no intra-racial/ethnic group inconsistencies between reports and ratings (e.g., worse reports but higher ratings).

The second major finding in this study is that language barriers may account for the lower ratings and worse reports given by Hispanics and Asians compared with whites. In all comparisons between whites and Hispanics, Hispanics who reported speaking English at home gave reports and ratings of care similar to whites, while Hispanics who reported
speaking Spanish at home gave lower reports and ratings than whites. Similarly, in all comparisons between whites and Asians, Asians who reported speaking English at home gave reports and ratings of care similar to whites, while Asians who reported speaking a language other than English at home gave lower reports and ratings than whites.

**Summary of Methodological Results**

*Chapter 3: Readability of CAHPS® 2.0 Child and Adult Core Surveys*

In Chapter 3, a readability analysis of the Adult and Child, English and Spanish versions of the CAHPS® 2.0 surveys is presented. Literacy level is an attribute of individuals, and refers to the reading ability level of an individual. Readability, on the other hand, is an attribute of written materials, and refers to the reading ability required to comprehend a text.

Concern about low literacy among potential respondents to self-administered surveys such as CAHPS®, makes this readability analysis pertinent. This concern is particularly acute regarding Medicaid respondents, who are increasingly being asked to respond to consumer surveys such as CAHPS®. According to the 1993 National Adult literacy survey (Kirsch, Jungeblur, Jenkins, & Kolstad, 1993), 75% of welfare recipients read at or below the eighth grade level and 50% read at or below the fifth grade level.

A mismatch between an intended respondent’s reading ability and the survey instrument may have important implications for the validity of patient evaluations research based on self-administered surveys. Some of the consequences of a mismatch may include low response rates,
especially in vulnerable populations, and unreliable responses because of poor item comprehension.

The readability analysis presented in this chapter is based on readability formulas. Readability formulas are mathematical formulas that predict the readability level of a text. Most readability formulas predict the readability level of a text from two measures: a measure of sentence length (syntactic variable) and a measure of word difficulty (semantic variable). These two variables are each multiplied by distinct constants and linearly combined with an intercept term to form a prediction rule for readability. Mathematically, a readability formula takes the form of:

\[ R = \alpha + \beta_1 S + \beta_2 W, \]

where \( R \) is the readability level, \( \alpha \) is the intercept term, \( \beta_1 \) is the syntactic parameter, and \( \beta_2 \) is the semantic parameter. The syntactic and semantic parameters in readability formulas are estimated based on studies of criterion passages of varying but known levels of reading difficulty.

For this study, five readability formulas were applied to the CAHPS® 2.0 surveys. The Fry Readability Graph is adapted for Spanish and English language documents. The FRASE Graph is applicable to Spanish language documents only. The Fog Index, SMOG Grading Formula, and the Flesch Reading Ease Score are only applicable to English language text.

Based on applying these formulas, the English versions of the CAHPS® 2.0 surveys had an estimated 6th to 8th grade readability level,
while the Spanish versions of the CAHPS® 2.0 surveys had an estimated 7th grade reading level. The readability estimates from this study are consistent with the opinion of an independent reading expert who assessed the readability level of the CAHPS® 1.0 surveys (Harris-Kojetin, Fowler, Brown, Schnaier, & Sweeny, 1999). They are also consistent with the observations of researchers who conducted cognitive interviews on the CAHPS® 1.0 survey instruments (Brown, Nederend, Hays, Short, & Farley, 1999).

Chapter 7: Evaluating the Equivalence of Health Care Ratings by Whites and Hispanics.

In Chapter 7, an assessment of the equivalence of ratings of care by Hispanic and non-Hispanic white survey respondents to the United Medical Group Association (UMGA) study survey is presented.

The purpose of this study is to assess nine rating questions asked in the survey for bias using statistical methods—more specifically item response theory (IRT) procedures. Bias, in the context of this study, refers to the observation that a question (item) displays different statistical properties in each of the two groups in the study—whites and Hispanics—after controlling for group differences in ratings. Put another way, this study seeks to identify items to which equally satisfied individuals from the different groups have unequal probabilities of answering in the same way. In psychometric parlance, this is a study of differential item functioning or DIF.

The survey data analyzed in this study are from the UMGA study, which is the same survey data analyzed in Chapter 4 and discussed in
section 2 of this chapter. Briefly, the survey data was collected from adult patients receiving care from the United Medical Group Association (UMGA), an association of physician groups located in the western United States. Of the 6,211 surveys analyzed in this study, Hispanics completed 713 and whites completed 5,508. The overall survey response rate was 59%.

Five items included in this study assess interpersonal aspects of care and four items assessed technical aspects of care. All nine items were administered using an identical 7-point response format (Very Poor, Poor, Fair, Good, Very Good, Excellent, and The Best) along with the option, Does Not Apply to Me. The five interpersonal care items asked survey respondents to rate: medical staff listening to what you have to say; answers to your questions; explanations about prescribed medications; explanations about medical tests and procedures; and reassurance and support from your doctor; and support staff. The four technical care items asked survey respondents to rate: quality of examinations; quality of treatments; thoroughness and accuracy of diagnosis; and comprehensiveness of exams.

Unidimensionality is assumed by the IRT models used in this study. Cronbach’s alpha of 0.96 was obtained for both whites and Hispanics. Principal components analysis was also conducted to test unidimensionality of the nine rating items. All principal components loadings were <0.83 for both the white and Hispanic groups. The ratio of the first to second eigenvalues was 17.8 and 17.3 for the white and Hispanic groups, respectively. The Tucker and Lewis coefficients for a one factor solution were 0.96 and 0.94 for the white and Hispanics groups, respectively. These results strongly suggest that there is a
single factor underlying the nine items and that the hypothesized nine-item scale is unidimensional. Previous studies have also demonstrated that evaluations of interpersonal and technical aspects of care may be represented by a single factor (Marshall, Hays, & Mazel, 1996).

The IRT analysis demonstrated that statistically significant DIF at the 0.05 level occurred in two of the nine items. The first DIF item was an interpersonal rating item and asked about reassurance and support; the second DIF item was a technical rating item and asked about quality of examinations.

To assess the clinical significance of these findings, additional analyses were conducted. First, the effect size of ethnicity – the standardized group mean difference – was assessed with and without inclusion of the DIF items in the ratings scale. Second, the test characteristic curves, based on all nine items, were computed for whites and Hispanics (test characteristic curves show the relationship between level of satisfaction and expected total test scores).

When all nine items were included in the scale, the effect size was 0.27, with whites rating care more positively than Hispanics \( (p<0.05) \). When the DIF items were dropped from the scale, the effect size was 0.26, with whites rating care more positively than Hispanics. Further, the test characteristic curves for whites and Hispanics are nearly identical by inspection (see Figure 1, page 277).

The results of this study show that although statistically significant DIF was detected in two of nine items used to measure patient evaluations of care, the amount of bias introduced by these items did not have a meaningful effect on a comparison of ratings
between whites and Hispanics. Thus the lower ratings of care by Hispanics should not be attributed to item bias, but rather to differences in actual experiences with care.

Chapter 8: Confirmatory Factor Analysis of the Consumer Assessment of Health Plans Study (CAHPS®) 1.0 Core Survey.

In Chapter 8, a confirmatory factor analysis of the CAHPS® 1.0 core survey is presented. The objectives of this study were:

- to confirm the hypothesized factor structure of the adult CAHPS® 1.0 core survey;

- to confirm the invariance of the adult CAHPS® 1.0 survey factor structure across white and Hispanic respondents, and commercial and Medicaid health sector respondents; and

- to examine the concordance between direct and indirect methods of summarizing consumer experiences with care.

The survey data analyzed in this study is the NCBD 1.0 data, which is the same survey data analyzed in Chapters 5 and 6 of this thesis and discussed in section 2 of this chapter. Briefly, the NCBD 1.0 is an aggregation of CAHPS® 1.0 survey results collected by health plans located across the United States. Health plans administering the CAHPS® surveys were asked to voluntarily contribute their survey results to the NCBD project for purposes of benchmarking and research. Although both adult and child survey results from Medicaid and commercial settings are included in NCBD 1.0, only the adult data were used in this study.
As in the previous study, only the core questions – 23 questions common to all versions of the surveys – were included in this study. These core evaluations can be divided into reports (indirect assessments) and ratings (direct assessments). Reports ask survey respondents to make judgments about the frequency of certain events. The CAHPS® 1.0 surveys include five reporting domains, each addressed by several questions.

The five reporting domains are access to needed care (four questions), provider communication (four questions), office staff helpfulness (three questions), promptness of needed care (four questions), and health plan customer service (two questions). With the exception of two questions which are asked using a Yes/No format, all questions in the reports domains are asked using a Never, Sometimes, Usually, Always format.

The CAHPS® 1.0 surveys also include four global rating questions. Each global rating item asks respondents to evaluate an aspect of care. The four rating items ask consumers to rate their personal medical provider, specialists, health care, and health plan. All four rating questions are administered using an 11-point numeric scale ranging from 0 to 10, anchored by statements like "The Best Possible Personal Doctor" and "The Worst Possible Personal Doctor."

Multiple group confirmatory factor analysis (CFA) of latent variables was used to address the objectives of this study. In confirmatory factor analysis, also known as structural equation modeling (SEM), hypotheses are translated into a series of mathematical equations that can be solved simultaneously to generate an estimated covariance matrix. By means of various goodness of fit indexes (e.g.,
CFI, NFI, NNFI), the estimated matrix can be evaluated against the observed covariance matrix to determine whether the hypothesized model is a good representation of the data. In general fit index values above 0.90 indicate acceptable model fit.

In the first set of analyses, the extent to which a common factor structure accurately characterized whites and Hispanics sub-samples in both Medicaid and commercial health sectors was tested. To address this objective, a four-group model, with the same patterns of item-factor relationships within each group but without any cross-group constraints, was estimated. These analyses confirmed that the five-factor data structure hypothesized in CAHPS® is well represented among all four subgroup; all goodness of fit indexes evaluated had values of greater 0.90. However, because of small differences in the valence of factor loadings between the health sectors, the subsequent analyses were conducted separately for the Medicaid and commercial subgroups.

In the second set of analyses, the extent of factorial invariance between whites and Hispanics was assessed. To test for weak factorial invariance, cross-group constraints were placed on analogous factor loadings between the white and Hispanic subgroups within each sector. Thus, not only were the patterns of item-factor relationships the same across ethnic groups as in the previous analyses, but the magnitudes and valances of those relationships were constrained to equality across groups. The results of these analyses showed that the model chi-square values increased to a statistically significant level (p<0.05) by adding the cross-group constraints in both the Medicaid and commercial sectors. However, by releasing 5 of 23 constraints in the commercial group, and 3 of 23 constraints in the Medicaid group, the change in
model chi-square did not reach statistically significance. These results indicate substantial but not complete weak factorial invariance between whites and Hispanics.

The third main set of analyses addressed the comparability of using direct and indirect evaluations of global evaluations of care. To address this question, a secondary factor, accounting for the covariance between the five reports factor, and a common primary factor, accounting for the covariance between the four rating questions, were hypothesized. To test the comparability of the direct and indirect measures, models with the correlation between the secondary factor (reports) and the primary factor (ratings) freely estimated and constrained to 1.0, were compared. In the unconstrained models, the correlation between the two factors was quite high in all groups (Hispanic/commercial: r=0.95; white/commercial: r=0.98; Hispanic/Medicaid: r=0.98; white/Medicaid: r=0.97). However, when the model chi-squared for the unconstrained and constrained were compared there were significant increases in both the Medicaid and commercial samples, indicating significant degradations in model fit. These results indicate that even though the indirect and direct approaches of measuring global satisfaction are highly correlated, they are not completely comparable, in a strict statistical sense.

**Implications for Health Policy**

This thesis provides evidence that significant racial/ethnic inequalities in patient’s evaluations of care exist. Specifically, the studies presented in Chapters 4, 5 and 6 show that Hispanics, Asians/Pacific Islanders, American Indians/Native Alaskans, persons of multiracial/other backgrounds, and persons who speak a language other
than English at home, report worse experiences with care and rate the care they receive lower than whites. Chapters 5 and 6 also show that although African American adults report better experiences with care and rate the care they receive higher that whites, African Americans report worse experiences with care and rate the care their child receive lower than whites. These findings suggest that efforts to improve the delivery of health services to racial and minorities, particularly those who encounter language barriers, are needed.

In a recent policy paper in the Journal of the American Medical Association, Fiscella et al. outlined five principles for reducing racial/ethnic and socioeconomic inequalities of healthcare quality (Fiscella, Franks, Gold, & Clancy, 2000). First, disparities in healthcare must be recognized as a significant quality problem. As the results of the research in this thesis shows and prior research has demonstrated, the healthcare system is not serving all members of society equally. Variation in quality of care generated by factors other than medical need or risk pose a critical challenge to quality in addition to raising questions of distributive justice.

The second principle proposed by Fiscella et al. (2000) is that health plans and other providers of health care need to collect relevant and reliable data to address racial/ethnic disparities. Health plans and other providers should collect racial/ethnic demographic data as part of the core data on patients. Without these data, efforts to illuminate racial/ethnic differences in quality of care are greatly hampered. Racial/ethnic group specific response rates cannot be computed. As a result, it is impossible to determine the representativeness of each sub-group in a survey sample.
Health plans and other providers need to ask about racial/ethnic demographic information on consumer surveys in a consistent manner. Although many health surveys now routinely collect this information, some do not. Further, standardized approaches for collecting this information are needed. Many surveys combine racial and ethnic backgrounds in a single question, forcing respondents to choose from a highly limited set of descriptors. More detailed information can be collected using two part questions. The 1999 current population survey questions about racial/ethnic background are a good example of this approach (www.bls.census.gov/cps/bqestair.htm).

Items about language preference and interpreters should also be included in health-related consumer surveys. Prior research (Woloshin, Bickell, Schwartz, Gany, & Welch, 1995a) as well as evidence presented in this thesis point to the importance of a patient’s primary language in determining his/her experiences with healthcare. Communication is a central feature of health care. Without good communication, optimal quality of care cannot be achieved (Woloshin et al., 1995a).

An optimal set of language items might include questions about language use and ability (Schneider, Riehl, Courte-Wienecke, Eddy, & Sennett, 1999), such as those in language acculturation scales (Marin, Vanoss Marin, Perez-Stable, & Vanoss Marin, 1990), and in the 1990 US Census (www.census.gov/population/www/socdemo/lang_use.htm).

In addition to defining a set of items for studying disparities, the items must produce reliable and valid data. Collecting reliable and valid data in multicultural settings is one of the central concerns of this thesis. Chapter 2 presents a framework for producing culturally and linguistically appropriate survey instruments. The
process begins by using focus groups and cognitive interviews to develop equivalent survey domains and items. Readability assessments are recommended to ensure that the survey instruments do not exceed the literacy level of the intended target audience. Finally, the instruments are tested for equivalence using psychometric methods such as item response theory or confirmatory factor analysis. In Chapters 7 and 8, methodological studies to evaluate the psychometric equivalence of two existing instruments are presented.

The third principle proposed by Fiscella et al. (2000) is that performance measures should be stratified by socioeconomic position and racial/ethnic background. Three of the studies presented in this thesis are studies of performance measurement stratified by racial/ethnic background. These studies have been prepared for publication in peer-reviewed journals. One is published and the other two are submitted. However, publication of these results in research journals is not sufficient. Despite concerns about the complexity of performance data (Hibbard, Slovic, & Jewett, 1997), stratified public reporting of health plan performance is needed. Until there is more complete public accountability of health plans for the care they deliver to their patients, it is unlikely that the particular needs of their diverse and vulnerable patients will be addressed adequately.

The fourth principle proposed by Fiscella et al. (2000) is that population wide performance measures be adjusted for socioeconomic position and racial/ethnic background. Fiscella et al. (2000) argue that because the racial/ethnic and socioeconomic backgrounds of enrollees are correlated with current performance measures, including measures of patient’s reports and ratings of care, adjustment would
facilitate more meaningful comparisons among health care providers. They further state that this step should not be undertaken until appropriate measures for monitoring care to vulnerable populations have been fully implemented to avoid institutionalizing substandard care.

The question of case-mix adjustment in general and whether to adjust comparisons among health plans for race/ethnicity specifically, has been a topic of research and discussion among the members of the CAHPS® research group at RAND (Elliott, Swartz, Adams, Spritzer, & Hays, submitted for publication). In principle, case-mix adjustment should only be used when a characteristic of survey respondents is correlated with a measure of quality of care, independent of the actual quality of care received.

Consider two alternative scenarios of race/ethnic differences in a measure of quality. In the first scenario, the racial/ethnic composition of two health plans differ because of a priori differential "assignment" to plans (or choice of plans) on the basis of race/ethnicity. This may occur because of geographic convenience, conscious choice, or other factors. Because minorities tend to rate lower than whites, the plan with the higher proportion of minorities will be rated lower. In the second scenario, pre-existing racial/ethnic differences do not exist, but plan performance in fact causes racial/ethnic difference to develop.

The first scenario constitutes the ideal situation for case-mix adjustment. Plan differences are due to the tendency of certain groups to give lower rating than others, independent of the care rendered by health plans. In the second scenario, case-mix adjustment for race/ethnicity is not advisable because it eliminates true information
on quality of care. In this case, adjustment for racial/ethnic differences prior to enrollment in the current plan is desirable, but adjustment for racial/ethnic differences since enrollment, or by extension racial/ethnic differences after enrollment in the current plan, is not desirable. In reality, probably a mixture of both scenarios exists. Thus, to the extent that the second scenario is true, case-mix adjustment for race/ethnicity is not advisable because race/ethnicity reflects true differences in quality rather than respondent bias. The methodological studies in Chapters 7 and 8 of this help determine the extent to which racial/ethnic differences in ratings are due to response bias.

The fifth and final principle Fiscella et al. (2000) suggest is that provider reimbursements include adjustments for the race/ethnicity and socioeconomic position of their patients. Presumably, these adjustments would account for the additional cost of providing care to minority patients and help offset the costs of quality improvement efforts designed to eliminate disparities.

**Implications for Future Research**

Future research efforts should focus on determining the appropriateness of case-mix adjustment for race-ethnicity and language use and preference. Although the two studies in this thesis address the issue of equivalence between whites and Hispanics, similar studies are needed to address equivalence among other racial/ethnic groups, including Asians, Pacific Islanders, African Americans, and American Indians. Further, cultural and linguistic differences between Asian and Hispanic subgroups may necessitate equivalence studies among these groups. For example, equivalence studies are needed between responded
answering to Spanish language versions of the CAHPS® survey and those answering to the English language version. Similar studies will be needed as the CAHPS® surveys are translated into other languages.

Although racial/ethnic differences are detectable using the current racial/ethnic categories in CAHPS® and UMGA, more refined racial/ethnic subgroup analysis is desirable. There is increasing evidence of substantial variation in access to care and satisfaction with care among Asians by country of origin (Murray-Garcia, Selby, Schmittdiel, Grumbach, & Quesenberry, 2000) and language spoken at home (see Chapter 6), and among Hispanics by language spoken at home (see Chapter 6). By aggregating up to major racial/ethnic group levels (e.g., Hispanics) important subgroup differences might be overlooked. In order to look at racial/ethnic subgroup, additional patient demographic variables will be needed. These might include for immigrants: country of origin, level of acculturation, and time spent in the United States; or for the US born: generational status and level of acculturation.

Finally, existing healthcare consumer surveys are far from perfect and require ongoing refinements. Both item revisions and changes in the domains may be needed as the health care system evolves. Future translations may necessitate item revisions or item deletion to maintain instrument equivalence. To address these and other pressing issues regarding the assessment of patient’s experiences with care through surveys, ongoing qualitative and quantitative research is needed.
References


Figure 1. Test characteristic curves for whites and Hispanics respondents.