RAND

Creating A Coordinated Autos/UAW Reporting System (CARS) For Evaluating Health Plan Performance

Elizabeth A. McGlynn, John Adams, Jennifer Hicks, David Klein

DRU-2123-FMC

September 1999

Prepared for DaimlerChrysler, Ford Motor Company, General Motors, and the United Auto Workers

DISTRIBUTION STATEMENT AUTHORIZATION RECORD

Title: Creating a Coordinated Autos/UAW Reporting System (CARS) for Evaluating Health Plan Performance

Authorizing Official: Earl Lightbody

Agency: RAND Ph. No. (310) 451-7002

☐ Internet Document: URL: _____
(DTIC-OCA Use Only)

Distribution Statement: (Authorized by the source above.)

☑ **A**: Approved for public release, distribution unlimited.

☐ **B**: U. S. Government agencies only. (Fill in reason and date applied). Other requests shall be referred to (Insert controlling office).

☐ **C**: U. S. Government agencies and their contractors. (Fill in reason and date applied). Other requests shall be referred to (Insert controlling office).

☐ **D**: DoD and DoD contractors only. (Fill in reason and date applied). Other requests shall be referred to (Insert controlling office).

☐ **E**: DoD components only. (Fill in reason and date applied). Other requests shall be referred to (Insert controlling office).

☐ **F**: Further dissemination only as directed by (Insert controlling DoD office and date), or higher authority.

☐ **X**: U. S. Government agencies and private individuals or enterprises eligible to obtain export-controlled technical data in accordance with DoD Directive 5230.25.

NOTES: _____

J Kingery
DTIC Point of Contact

1-5-2000
Date

TABLE OF CONTENTS

TABLES

FIGURES

CHAPTER 1
SETTING THE CONTEXT

Private purchasers of health insurance have been one of the key forces in stimulating the development of systematic methods to evaluate the quality of care provided in managed care. There are at least two reasons for purchasers' interest in quality. First, as employers began selecting health insurance options, such as managed care organizations, that promised to reduce the rate of increase in premiums, employees' concerns expressed that the quality of care would decline. The fear among consumers was that limits on choice and financial incentives to limit utilization, the hallmarks of early managed care organizations, would translate into poorer quality. Second, as the cost of providing health insurance began to consume larger portions of the budgets for producing goods and services, employers began to view health insurance as an "input" much like windshields, tires, or electrical systems. Purchasers have become accustomed to assessing, monitoring, and continuing to improve the quality of other inputs to production processes, so viewing health through a new lens was an obvious next step. At the same time, many of the techniques for creating improvement in industrial processes began being adapted for application in the health sector, suggesting that improved monitoring could stimulate better service delivery.

Until recently, little information has been systematically available on the quality of care delivered in the U.S. Most of what is known comes from research studies that focus on a narrow range of conditions in a small number of settings (Schuster, McGlynn, Brook, 199__). The National Committee for Quality Assurance (NCQA) responded to demands for better information by facilitating a collaboration among purchasers, managed care plans, and consumers to develop a method, the Health Plan Employer Data and Information Set (HEDIS) to systematically collect, analyze and report data on health plan performance. NCQA also accredits managed care organizations. These efforts have substantially increased the amount of information available on health plan performance, and have led to the next problem--making sense out of the information.

In the Spring of 1998, the "Big Three" automobile manufacturers—DaimlerChrysler, Ford, and General Motors—and the United Auto Workers (UAW) agreed to collaborate on developing a method for summarizing the performance of managed care plans. All four had previously undertaken such efforts, but because the methods used by each one were different, the results also varied. For example, in the 1997 open enrollment materials for Ford and GM, three of the performance categories were similar: prevention or effectiveness of care, access to care, and consumer satisfaction. But an examination of the "scores" for 40 plans with which both

companies contracted revealed agreement on ratings for less than half of the plans. Because many households in southeast Michigan include family members who work for more than one automobile manufacturer, the potential for confusion resulting from these apparent mixed messages is significant. Such confusion is likely to undermine the shared vision of all four organizations that information on quality will help consumers make better choices. The UAW played a key role in facilitating this collaboration because its members work for all three automobile manufacturers, making the importance of consistent messages to its membership essential. The automobile manufacturers and the UAW have a history of collaboration on providing information to consumers to facilitate decision making and this project was a natural outgrowth of these prior efforts.

A team at RAND was selected to develop the method for reporting on health plan performance under the direction of the Coordinated Autos/UAW Reporting System (CARS) Steering Committee consisting of representatives from each of the three automobile manufacturers, the United Auto Workers (UAW), the State of Michigan (both as an employer and as a purchaser of services for Medicaid enrollees), and the Greater Detroit Area Health Council (GDAHC). In addition, RAND collaborated with the National Committee for Quality Assurance (NCQA) and the Foundation for Accountability (FACCT) through a technical advisory committee. RAND, NCQA, and FACCT have worked together on other quality measurement and reporting projects. Over the four and one-half months of the project, the CARS Steering Committee met four times to discuss policy questions and provide direction and feedback on the project.

The purpose of this report is to describe the method that was developed and the rationale for the choices that were made. The subsequent chapters discuss creating a framework for reporting on performance (Chapter 2), the sources of data on performance that were considered and eventually selected (Chapter 3), the method by which individual measures were assigned to categories (Chapter 4), methods for aggregating individual measures into summary scores (Chapter 5), choosing the reporting strategy (Chapter 6), obtaining data from health plans (Chapter 7), the final results that were reported to the CARS Steering Committee on August 14, 1998 (Chapter 8), and a discussion of the project with recommendations for next steps (Chapter 9).

CHAPTER 2
DEVELOPING A FRAMEWORK FOR REPORTING ON PERFORMANCE

WHY IS A FRAMEWORK USEFUL?

One of the motivations for the current project, as well as numerous efforts in the popular press (e.g., U.S. News and World Report, Newsweek, Consumer Reports), was to summarize a variety of individual measures about health plan performance into a few dimensions. Why is this necessary? Even the brightest human being can only hold a few pieces of information in short term memory when making a decision. Cognitive psychologists suggest that about five to seven bits of data can be utilized when making a decision. Further, hierarchical structures that organize specific details within a general framework facilitate the use of information in three ways. First, hierarchies facilitate comprehension. Second, hierarchies help people memorize information and retrieve that information for later use. Third, hierarchies communicate importance. The framework used for the CARS results, thus, should have a few categories and should organize information in a way that is useful for decision makers.

WHAT APPROACHES CAN BE USED TO CREATE A FRAMEWORK?

There are two different strategies for creating frameworks. The first approach, which might be called "bottom-up," starts with the individual measures that are available and creates summary categories that maximize the number of measures used. This can either be done quantitatively, using factor analysis or other methods designed to identify patterns in data, or it can be done qualitatively by obtaining expert opinion. The second approach, which might be called "top-down," starts with the information that potential users would like to have to make decisions and identifies measures that communicate the desired information. The methods for identifying what information the target audience wants may include surveys, focus groups, or semi-structured interviews.

The bottom-up approach is more frequently associated with research or decision analysis. This approach has the advantage of trying to use all available information. Since the approach is empirically driven, another advantage is the opportunity to identify patterns in data that might otherwise have escaped notice. The disadvantage of this approach, particularly if done quantitatively (e.g., using factor analysis), is that it may produce results that are difficult to interpret and may not be valued by the intended audience.

The top-down approach is more audience sensitive because it identifies attributes that are important to those making the decision. Because decision-makers generally come to a task with some questions already in mind, an optimal top-down approach organizes information into categories that respond to the questions on the minds of potential users. The disadvantage of this

approach is that there may be categories of interest to decision makers for which no or few measures currently exist.

HOW DO THESE TWO APPROACHES COMPARE?

Four existing frameworks illustrate the choices that were available when this project was started: Ford 1997, U.S. News and World Report, FACCT, and NCQA. The methods by which each framework was developed and the resulting categories are discussed in this section.

Ford 1997

In 1997, the Ford Motor Company contracted with RAND to develop a method for summarizing the available performance information for both the managers choosing plans to offer Ford employees, and for its employees to use in making decisions during open enrollment. Ford was interested in maximizing the use of information, so a bottom-up strategy was employed. The RAND team examined the available measures from HEDIS®, NCQA's Member Satisfaction Survey, NCQA Accreditation, GeoAccess, and the Ford Direct Questionnaire and recommended that summary scores be developed on performance in six categories: effectiveness of care, consumer assessments of care, organizational structure, targeted intervention programs, and resource utilization. These are described below.

Effectiveness of Care. For a measure to be included in the effectiveness of care domain, there must be scientific evidence that greater adherence to the process being measured will result in improved health for the population served or that the outcome measured is likely to be substantially influenced by actions taken by the plan or providers. Finally, it is reasonable to conclude that variations in performance on the measures in this category can be attributed to differences in the quality of service delivery in the health plan.

The following HEDIS 3.0 (1997) effectiveness of care measures populated this category: advising smokers to quit, breast cancer screening, cervical cancer screening, beta blocker treatment after heart attack, childhood immunization rate, adolescent immunization rate, prenatal care in the first trimester, check-ups after delivery, and eye exams for persons with diabetes.

Access to Care. Measures of access to care reflect those factors that have been shown to either facilitate or inhibit an individual's ability to obtain needed services. While these measures may not always be directly interpretable as meaning that individuals served by a plan lack access to care, they flag potential areas of concern. Access to care measures include both objective measures (e.g., capacity of the physician network, proximity of service locations to the population) and subjective measures (e.g., ratings by customers of experiences in trying to obtain services).

Measures in this category came from the Member Satisfaction Survey, Ford Direct Questionnaire, HEDIS access/availability of care measures, and GeoAccess data. Four

subcategories were developed: new member access to physicians, organizational characteristics, consumer ratings of access, and distance to providers.

Consumer Assessments of Care. Measures in this category reflect consumers' ratings of their experiences with various aspects of the health plan delivery system. Standardized survey tools facilitate comparisons among health plans in multiple areas including interactions with primary care and specialty physicians, financial aspects, and overall ratings of the health plan.

Measures in this category came from the NCQA Member Satisfaction Survey and the Direct Questionnaire. Four subcategories of measures were included: doctors and medical care, health plan cost and coverage, health plan information and administration, and health plan assessments.

Organizational Structure. Measures in this category reflect the capacity of the plan to manage care for the enrolled population. A major component of this assessment is the accreditation status of the plan, but information about the quality of physicians in the plan, use of various management tools, financial and personnel stability, and responsiveness to requests from Ford were also included.

Measures in this category came from NCQA Accreditation, URAC Accreditation, Ford Direct Questionnaire, and HEDIS. Five subcategories were created: accreditation, physician panel, plan stability, management programs, and responsiveness to Ford.

Targeted Intervention Programs. Measures in this category evaluate the extent to which the health plan has special programs that address the health needs of individuals with chronic diseases, multiple conditions, as well as health promotion programs. A particular emphasis for this category was on programs for persons with cardiovascular disease, asthma, diabetes, and mental health diagnoses.

Measures in this category came from HEDIS and the Ford Direct Questionnaire. Six subcategories were created: cardiovascular/cerebrovascular, other chronic conditions, prenatal care programs, prevention programs, health status assessment, and mental health/chemical dependency.

Resource Utilization. Measures in this category provide insight into the relative intensity of resource utilization among different health plans. Information about the use of particular procedures, types of admissions, and use of care in various settings are summarized by the proportion of premium dollars that are accounted for by an estimated expenditure on selected categories of utilization.

Measures in this category came from HEDIS and the Ford Direct Questionnaire. Seven subcategories were created: frequency of selected procedures, high occurrence/high cost DRGs, acute inpatient services, nonacute inpatient services, ambulatory care, mental health/chemical dependency, and outpatient drug.

U.S. News and World Report

The National Opinion Research Center (NORC), under contract to U.S. News and World Report (USNWR), used a quantitatively-based, bottom-up methodology to create summary categories of HMO performance. The USNWR/NORC method relied exclusively on HEDIS and Member Satisfaction Survey data. Factor analytic techniques were used to develop the categories.[1] Five clusters of 28 measures explained two-thirds of the variation in plan scores.

Prevention. This category included 11 measures: childhood immunizations, well-child visits in the first 15 months of life, well child visits for ages 3-6, well adolescent visits for ages 12-21, prenatal care in the first trimester, check-ups after delivery, breast cancer screening, cervical cancer screening, beta blocker treatment after a heart attack, retinal exams for persons with diabetes, and follow-up after hospitalization for a mental illness.

Access to Care for Adults. This category included the percent of adults in three age groups (20-44, 45-64, 65 and older) who had a preventive or ambulatory care visit in the previous two years and the percentage of primary care providers in the plan accepting new patients.

Member Satisfaction. This category included the percentages of enrollees responding positively to questions on NCQA's Member Satisfaction Survey about overall satisfaction, receiving needed care, ability to obtain referrals to specialists, choice of physicians, and the ability to make appointments.

Physicians' Credentials. This category included the percentages of primary care providers, physician specialists, pediatricians, OB/GYNs, and geriatricians who are board certified.

Access to Care for Children. This included the proportion of children in three age groups (12-24 months, 25 months to 6 years, 7-11 years) that had a visit with a primary care provider in the previous year.

None of the use of services measures related to procedures, specific DRGs, mental health/chemical dependency, inpatient, or outpatient drug were used in the USNWR/NORC model. This method was able to produce scores for 271 plans in 45 states using a four star reporting strategy; 18 plans were excluded due to insufficient data.

Foundation for Accountability (FACCT)

In contrast to the expert analytic approaches described above, the Foundation for Accountability (FACCT) has developed a top-down Consumer Information Framework. The model was developed using a focus group methodology that elicited from consumers the type of

---

[1] More detail about the scoring methodology can be found on NORC's Web page: http://www.norc.uchicago.edu/new/hmo.htm

information they need to have available in order to choose among health plans or providers. This approach both creates categories that reflect how the target audience thinks about the key issues and labels those categories with terms that are meaningful to the audience. The model has five components that are described below.

The Basics. Measures in this category describe how well health plans deliver the basics of good care--access, skill, communication, coordination of care and follow-up.

Staying Healthy. Measures in this category describe how well health plans help people avoid illness and maintain health through education, prevention and risk reduction.

Getting Better. Measures in this category describe how well health plans help people recover when they're sick or injured with appropriate treatment and follow-up.

Living With Illness. Measures in this category describe how well health plans help people with chronic conditions reduce symptoms, avoid complications and maintain daily activities.

Changing Needs. Measures in this category describe how well health plans care for people and their families at the end of life or when functional abilities change dramatically.

FACCT has mapped measures from a variety of current and potential measure sets (e.g., HEDIS, CAHPS, FACCT, ORYX) into this framework and has tested the placement of those individual measures with additional focus groups.[2]

National Committee for Quality Assurance (NCQA)

NCQA collaborated with FACCT in the development of a framework that could be used to report accreditation results by category rather than a single result. Additional focus groups were conducted with both consumers and purchasers to develop these categories which closely resemble those originally developed by FACCT. The system has five categories as well, but one of FACCT's original categories (The Basics) is split into two and one category (Changing Needs) is not included.

Access and Service. Measures in this category reflect experiences of enrollees in choosing doctors, making appointments, obtaining desired care, getting specialty referrals, and resolving problems. Additional measures may include objective assessments during accreditation of the plan's ability to provide access to needed services.

Qualified Providers. Measures in this category reflect the quality of individual doctors and medical groups under contract to the health plan. This may include consumer ratings as well

---

[2] More information about FACCT's framework and measurement work can be found on FACCT's Web page: http://www.facct.org.

as objective measures, such as the proportion of doctors in different specialties who are board certified.

Staying Healthy. Measures in this category reflect how well the health plan helps people avoid illness through preventive care, reduction in health risks and early detection of serious disease.

Getting Better. Measures in this category reflect how well the health plan helps people recover when they are sick or injured.

Living with Illness. Measures in this category reflect how well the health plan helps people with chronic conditions maintain or improve the quality of their lives and avoid complications of serious illnesses.

NCQA is transitioning to using this framework for reporting accreditation and HEDIS performance measures. Each category would include information from the accreditation process as well as a subset of the HEDIS performance measures. Benchmarks are established that incorporate performance within the region the health plan operates as well as national performance.[3]

Comparison of Categories from Different Methods

Despite the different approaches to constructing frameworks, there are some remarkable consistencies among the bottom-up and top-down approaches as illustrated Table 2.1.

[3] Accreditation results, HEDIS highlights, and benchmarks can be found on NCQA's Web page: http://www.ncqa.org.

Table 2.1

Comparison of Summary Categories

Ford 1997	USNWR	FACCT	NCQA
Organizational Structure	Physician's Credentials	The Basics	Qualified Providers
Access to Care	Adult's Access Children's Access	The Basics	Access and Service
Consumer Assessments	Member Satisfaction		
Effectiveness of Care: Prevention	Prevention	Staying Healthy	Staying Healthy
Effectiveness of Care: Acute		Getting Better	Getting Better
Effectiveness of Care: Chronic and Targeted Intervention Programs		Living with Illness	Living with Illness
Resource Utilization			

WHAT WERE THE FINAL CARS CATEGORIES?

The Steering Committee for the Coordinated Autos/UAW Reporting System (CARS) project developed a set of categories that are closely related to those used by FACCT and NCQA but that reflect some priorities of concern to the Steering Committee representatives. Five categories were included in the final system.

NCQA Accreditation Status. Based on a review by an independent group of health professionals, health plans are evaluated against quality standards. Site visits are included in the accreditation determination.

Consumer Satisfaction. This category provides a summary score of the overall satisfaction of HMO members with their health plan, including how much they were helped by the care received, whether or not they would recommend the plan to family or friends, and whether they plan to remain enrolled.

Access and Service. Measures in this category reflect experiences of enrollees in choosing doctors, making appointments, obtaining desired care, getting specialty referrals, and resolving problems. Additional measures may include objective assessments during accreditation of the plan's ability to provide access to needed services.

Staying Healthy. Measures in this category reflect how well the health plan helps people avoid illness through preventive care, reduction in health risks and early detection of serious disease.

Getting Better/Living with Illness. Measures in this combined category reflect how well the health plan helps people recover when they are sick or injured and how it helps people with chronic conditions maintain or improve the quality of their lives and avoid complications of serious illnesses.

The CARS choices reflect both an interest in maintaining consistency across reporting years and a desire to produce information that employees will find useful. The Big Three and the UAW have placed considerable emphasis on NCQA Accreditation Status as a minimum condition for plans with which they contract. The accreditation status had been reported to consumers in previous years and there was an interest in maintaining the importance of this quality signal. Although there is relatively little variation in this measure, because most of the plans that contract with the automobile companies are already accredited or are seeking accreditation, the continued importance of this objective assessment of health plan quality was reflected in the decision to report the category.

While the Steering Committee was supportive of NCQA's decision to divide FACCT's The Basics category into two, the category "Qualified Providers" received little support from the group. In particular, the Steering Committee was concerned that plans scoring low on this category might be interpreted to have providers that were not qualified and consumers might wonder why such plans were being offered. These concerns were only increased when the few measures that could be used to score this category were examined. These measures were felt to be too narrow in scope to reflect an evaluation of provider qualifications.

Neither the FACCT nor NCQA frameworks include a category called Consumer Satisfaction, although both Ford 1997 and USNWR included such categories (Consumer Assessments of Care and Member Satisfaction, respectively). In previous years, The Big Three had reported on Consumer Satisfaction with health plans. Given the push to enroll employees in managed care, the Steering Committee was concerned that if Consumer Satisfaction was dropped as a category, employees would interpret this as a signal that satisfaction was no longer a priority for purchasing decisions.

The decision to merge Getting Better and Living with Illness was made after determining that only four measures were available to score those categories (two measures in each category). The CARS Steering Committee believed that this was too few to support a category score. So, the two categories were combined. The intent is to split these categories apart in the future as more measures become available. A number of new chronic disease measures are on the horizon to be added to HEDIS, so this division may occur within the next couple of years.

CHAPTER 3
SOURCES OF DATA

Health plan performance information is available from a variety of sources. The following data sources were evaluated for inclusion in the CARS model: The Health Plan Employer Data and Information Set (HEDIS), NCQA's Member Satisfaction Survey (MSS), The Consumer Assessment of Health Plans Survey (CAHPS), accreditation status (NCQA, URAC, JCAHO), GDAHC/GM/DaimlerChrysler's Request for Information (RFI), Ford's Direct Questionnaire, GeoAccess, and The Foundation for Accountability (FACCT).

DESCRIPTION OF DATA SOURCES

The Health Plan Employer Data and Information Set (HEDIS)

HEDIS is a standardized set of health plan performance measures that are collected and maintained by the National Committee for Quality Assurance (NCQA). HEDIS measures evaluate the effectiveness, availability, costs, and utilization of health care services within a managed care plan. Approximately 300 of the nation's managed care plans submit HEDIS data to NCQA on an annual basis for public release. HEDIS 3.0/1998 (the measure set reported in 1998 on performance in calendar year 1997) contains 46 measures plus a member satisfaction survey for commercial enrollees in health plans. Many of the plans that do not submit HEDIS performance data to NCQA for public reporting do collect the data for internal use or to meet contracting requirements.

NCQA's Member Satisfaction Survey (MSS)

The MSS is a standardized survey that addresses satisfaction with care and is a required component of HEDIS 3.0/1998. The MSS contains 79 items that examine the following content areas: socio-demographic characteristics of the respondent, screening questions for length of coverage by the health plan, satisfaction with care and the health plan's services, and the respondent's health and daily activities.

The Consumer Assessment of Health Plans Survey (CAHPS)

CAHPS is a standardized satisfaction survey that is completed by a health plan's members. The Adult Core Questionnaire from CAHPS contains 46 items; nine additional measures appear in CAHPS' Child Core Questionnaire. The CAHPS and HEDIS satisfaction surveys will be converged for calendar year 1998 and reported in HEDIS 1999. A converged survey was not available in HEDIS 3.0/1998.

Accreditation Status (NCQA, URAC, JCAHO)

NCQA offers accreditation to managed care organizations. Over 300 health plans nationally have been or are scheduled to be reviewed by NCQA for accreditation. The American Accreditation Healthcare Commission/URAC offers eight different accreditation programs for managed care plans, although the organization's primary focus has been on accreditation of preferred provider organizations (PPOs). The Joint Commission on Accreditation of Healthcare Organizations (JCAHO) offers accreditation to hospitals, health care networks, PPOs, and health care organizations that provide home care, long term care, behavioral health care, and laboratory and ambulatory services.

Request for Information (RFI)

DaimlerChrysler and General Motors, in conjunction with the Greater Detroit Area Health Council (GDAHC), require health plans with which they contract to participate in an RFI process. The 1998 RFI consists of over 400 questions that collect information regarding the financial, clinical, and service performance of the health plan. A contractor then conducts site visits to supplement the assessment.

Ford's Direct Questionnaire

Health plans that contract with Ford are required to fill out a Direct Questionnaire. The Direct Questionnaire contains a subset of the RFI questions that collect information on health plans' disease management and prevention programs.

GeoAccess

GeoAccess is a proprietary system that provides an objective measure of member's geographic proximity to health care providers and hospitals. GeoAccess can be used to identify the proportion of beneficiaries that are within an "ideal" or "acceptable" drive to one or more primary care physicians and participating hospitals.

Foundation for Accountability (FACCT)

FACCT is a not-for-profit organization that was established to identify, develop and endorse measures of health care performance. In addition to its consumer information framework, FACCT has proposed a number of performance measures designed to capture consumer-relevant information.

CRITERIA FOR SELECTING DATA SOURCES

After the potential data sources for health plan performance data were identified, they were evaluated on four criteria: availability, timeliness, reliability, and the ability to distinguish

between plans. Table 3.1 displays a mapping of the measures in each of the data sources to the criteria; an "X" means that a majority of measures in the data source meet the criterion.

Availability

Inclusion of a data source depended on how likely it was to be available on most health plans with which the three automobile manufacturers contract. A coordinated strategy for health plan performance reporting requires that a common set of performance measures be used. Measures that were not available for the majority of health plans with which the automobile manufacturers contract could not be used to effectively compare performance across plans.

Timeliness

A data source was only included if the results were expected to be available in time to produce results for GM and DaimlerChrysler's fall open enrollment period (Ford has a winter open enrollment period). In order to score performance and provide the results to GM and DaimlerChrysler for the publication of their open-enrollment materials, data needed to be available to RAND by the end of July, 1998. Measures could meet the timeliness criterion without meeting the availability criterion (i.e., if results could be obtained within the timeframe for only a subset of plans).

Reliability

The data sources also needed to be a reliable source of information. The intent of providing a coordinated strategy for reporting on quality was to send a consistent and accurate message about health plan performance to the marketplace. Data from unreliable sources would have the potential to misrepresent the true performance of health plans. Representatives from the automobile manufacturers and the UAW were strongly committed to communicating an accurate message on health plan performance to the marketplace and specifically to their beneficiaries.

Distinguish Among Plans

Finally, a data source needed to be useful in distinguishing among health plans. To identify differences in quality, measures are most useful when there is variation in performance or when performance is consistently poor. We evaluated whether plans were likely to have significantly different scores on the measures in each source.

Table 3.1

Data Sources and Inclusion Criteria

	Criteria for Evaluating Data Sources			
Data Source	**Available**	**Timely**	**Reliable**	**Distinguishable**
HEDIS	X	X*	X	X
MSS	X	X*	X	X
CAHPS			X	X
NCQA Accreditation	X	X	X	X
URAC Accreditation		X	X	X
JCAHO Accreditation		X	X	X
RFI	X**	X	?	?
Direct Questionnaire	X**	X	?	?
GeoAccess		X	X	
FACCT			?	?

* The HEDIS and MSS measures were available directly from the health plans in time to be used for the open enrollment materials. However, HEDIS and MSS measures were not available via a direct data feed from NCQA to meet the open enrollment deadline (see Chapter 7 on Obtaining Data from Health Plans).

** Availability of RFI and Direct Questionnaire measures for all health plans refers only to the common set of disease management and prevention measures found in both data sources.

INITIAL SELECTION OF DATA SOURCES

Given the four criteria for the inclusion of a data source (e.g., available, timely, reliable, and distinguishable), the initial selection of data sources included the following: HEDIS, MSS, NCQA Accreditation, and the common set of measures found in the RFI and Direct Questionnaire.

CAHPS was eliminated as a data source for 1998 performance scores because the survey was not fielded on enrollees in most health plans. Subsequent iterations of the CARS model will incorporate the converged CAHPS and HEDIS satisfaction surveys.

Accreditation status is believed to be a reliable way to distinguish health plan quality. Due to contract requirements, there is little variation in accreditation status among the health plans with which the Big Three contract. However, in the broader universe of health plans, accreditation status is a significant indicator for health plan quality. All three of the accrediting bodies post the accreditation status of reviewed organizations on their Web pages. The availability of URAC and JCAHO Accreditation status was extremely limited for the HMOs that were to be evaluated under the CARS model, consequently they were excluded from the model. NCQA Accreditation was maintained as a data source because it met all of the defined criteria, including availability for the majority of plans.

The overlapping set of disease management and prevention program questions were included in the initial selection of data sources. During the initial selection process it was clear

that the data from the RFI and Direct Questionnaire would be available for the health plans with which the autos contract on a timely basis. However, little was known about the other characteristics (i.e., reliability and ability to distinguish between health plans) of the data because this was the first time that DaimlerChrysler, Ford, GM and GDAHC had coordinated to incorporate identical questions for disease management and prevention programs. The RFI/Direct Questionnaire measures for disease management and prevention were eventually dropped from the scoring strategy in large part because health plans felt very strongly that the data were highly unreliable.

Ford had previously used information from GeoAccess to assess their beneficiaries' access to primary care physicians and hospitals. GeoAccess scores must be purchased for each health plan, so while the data were available, an extra expenditure would have been required. GeoAccess results demonstrated little variation in access to providers between the plans with which Ford contracts. It was decided that the information gained through GeoAccess best informs the decision of whether or not to contract with a health plan. Because the purpose of the CARS project was to provide information to consumers for choice among plans already selected by the CARS, GeoAccess was excluded from the system based on its relatively poor discrimination among plans.

CHAPTER 4
ASSIGNING MEASURES TO CATEGORIES

Once a framework and data sources had been selected, the next task was to assign individual measures (for surveys, these are often called items) to the appropriate framework category. There are three basic approaches to this task:

- Expert judgment: A set of experts who understand the content of the measures to be included in the system use the framework category definition to guide assignment of measures to categories. Generally, this is done with multiple experts and formal or informal methods may be used to arrive at consensus.
- Consumer judgment: One or more groups of consumers could be asked to assign measures to categories. This may be done in focus groups, one-on-one interviews, or through a large-scale voting method.
- Analytic assignment: Methods such as factor analysis can be used to assign measures to categories based on empirically observed relationships among the measures. The USNWR/NORC method described in Chapter 2 is an example of how categories were created and measures assigned using factor analytic techniques.

Each of the methods has advantages and disadvantages. In the interest of time and budgets, the CARS Steering Committee selected the first approach, expert judgment, to assign measures to categories.

The method that was used could be called a "very modified Delphi method." The first step was for RAND to make an initial selection of measures for inclusion from each of the data sources that had been selected. For example, since the CARS framework does not include categories that make use of cost or utilization information, most of the measures in the HEDIS use of services category were not included.

Representatives from RAND, FACCT and NCQA were asked to assign the selected measures to the CARS framework categories independently. Because of the short timeframe, formal votes were received from only RAND and FACCT. Representatives from each of the three organizations then met to discuss differences and to resolve as many as possible. Most differences were satisfactorily resolved.

RAND presented the results of the expert process to the CARS Steering Committee which was asked to evaluate the results and resolve any outstanding differences.

The results of this process are summarized in Table 4.1. The first column lists the individual measures (which for the survey are individual questions). The next six columns

reflect the votes made by RAND and FACCT by category, including additional recommendations to exclude the measure from the framework. The seventh column contains a flag for disagreements. The last column shows the final category assignment; this process resulted in 43 measures being assigned to categories.

The voting was done before the framework had been completely decided, so this table reflects the NCQA framework rather than the final CARS framework. The information is presented to give an idea of the areas of controversy and to illustrate the method that was used. Expert opinion was also used to create the Consumer Satisfaction category that ultimately was included in the CARS framework.

Two examples of how viewpoints can affect assignments are presented to illustrate some of the differences we encountered. One of the HEDIS measures is Breast Cancer Screening, which reflects a plan's mammography screening rate for women age 50 and older. From a clinical perspective, mammography is used to identify breast cancer at a sufficiently early stage that the opportunities for a cure are substantially increased. This would imply that the measure could be placed in the Getting Better category (helping people who are sick recover). From a consumer perspective, however, mammography is done to help people avoid illness and as such belongs in the Staying Healthy category. For CARS, breast cancer screening was placed in the Staying Healthy category.

Another measure, obtained from survey data, determines whether persons in a plan who smoke were advised by their physician to stop smoking. From a consumer perspective, this measure reflects how well health plans help people avoid illness and reduce risks. From a clinical perspective, many smokers are already suffering health consequences from their behavior which would place the measure in the Getting Better category. One good time to get people to stop smoking is when they are suffering from other illnesses, particularly respiratory or cardiac problems. This measure was ultimately placed in the Getting Better category.

Analytic methods, such as examining correlations between items and scales, can also be used to evaluate where a measure should be placed. Such methods were not utilized during this round.

Table 4.1
Results of Expert Voting to Assign Measures to Categories

	ORIGINAL VOTES (FACCT and RAND) BY CATEGORY								FINAL CATEGORY ASSIGNMENT
MEASURE	Access and Service	Qualified Providers	Staying Healthy	Getting Better	Living with Illness	Exclude from Aggregation Strategy		Disagree? (1=yes)	1= NCQA Accred; 2= Access & Serv; 3 = Cons. Sat; 4 = Stay Healthy; 5 = GB/L w/ Ill; 9 = drop
NCQA Accreditation	1	0	0	0	0	1		1	1
7. Thinking about your own health care and the services you received from [Health Plan Name] over the last 12 months, how would you rate the following? a. Ease of making appointments for medical care by phone	2	0	0	0	0	0		0	2
7b. Length of time you had to wait between making an appointment for routine care and the day of your visit	2	0	0	0	0	0		0	2
7c. Thoroughness of treatment you received	0	2	0	0	0	0		0	2
7d. Attention given to what you had to say	0	2	0	0	0	0		0	2
7e. Number of doctors you had to choose	2	0	0	0	0	0		0	2
7f. Ease of choosing a personal physician	2	0	0	0	0	0		0	2
8. Thinking about your own health care and the services you received from [Health Plan Name] over the last 12 months, how would you rate the following? a. Types of services the plan covered	2	0	0	0	0	0		0	2
8b. Availability of information from your plan about eligibility, covered services or administrative issues	2	0	0	0	0	0		0	2
8c. Availability of information from your doctor or plan about costs of care	2	0	0	0	0	0		0	2
8d. Length of time you had to spend filling out claim forms or other paperwork	2	0	0	0	0	0		0	2
9. Have any of the following been a problem for you in arranging for your medical care in the last 12 months? If so, how much of a problem? a. Delays in your medical care while you waited for approval by your health plan	2	0	0	0	0	0		0	2
9b. Difficulty in receiving care you and your doctor believed was necessary	2	0	0	0	0	0		0	2
9c. Not being able to get a referral to a specialist that you wanted to see.	2	0	0	0	0	0		0	2
15. Over the last 12 months, how many days did you usually have to wait between the time you made an appointment for care and the day you actually saw the provider? a. Routine care (like a check-up)	2	0	0	0	0	0		0	2
15b. Minor illness or injury (like treatment for a sore throat)	2	0	0	0	0	0		0	2
16. Over the last 12 months, when calling for medical information or advice, how long did it usually take for your provider's office to return your call?	2	0	0	0	0	0		0	2
17. Over the last 12 months, once you got to your provider's office, how long did you usually have to wait to see your provider when you had an appointment for care?	2	0	0	0	0	0		0	2
18. Over the last 12 months, when you went for medical care, how often did you see the same provider?	2	0	0	0	0	0		0	2
7. Thinking about your own health care and the services you received from [Health Plan Name] over the last 12 months, how would you rate the following? g. Amount of time you had with doctors and staff during visit	1	1	0	0	0	0		1	2
10. Have you called or written [Health Plan Name] with a complaint or problem in the last 12 months?	1	0	0	0	0	1		1	2
10a. How long did it take for the health plan to resolve your complaint?	1	0	0	0	0	1		1	2
15. Over the last 12 months, how many days did you usually have to wait between the time you made an appointment for care and the day you actually saw the provider? c. Chronic or ongoing condition	1	0	0	1	0	0		1	2

Table 4.1
Results of Expert Voting to Assign Measures to Categories

MEASURE	ORIGINAL VOTES (FACCT and RAND) BY CATEGORY							FINAL CATEGORY ASSIGNMENT
	Access and Service	Qualified Providers	Staying Healthy	Getting Better	Living with Illness	Exclude from Aggregation Strategy	Disagree? (1=yes)	1= NCQA Accred; 2= Access & Serv; 3 = Cons. Sat; 4 = Stay Healthy; 5 = GB/L w/ Ill; 9 = drop
15. Over the last 12 months, how many days did you usually have to wait between the time you made an appointment for care and the day you actually saw the provider? d. Urgent care	1	0	0	1	0	0	1	2
7. Thinking about your own health care and the services you received from [Health Plan Name] over the last 12 months, how would you rate the following? h. How much you were helped by the care you received	0	2	0	0	0	0	0	3
11. All things considered, how satisfied are you with your current health plan?	2	0	0	0	0	0	0	3
13. Would you recommend your current health plan to your family or friends if they needed care?	2	0	0	0	0	0	0	3
14. Do you intend to switch to a different health plan when you next have an opportunity?	2	0	0	0	0	0	0	3
7. Thinking about your own health care and the services you received from [Health Plan Name] over the last 12 months, how would you rate the following? i. Overall quality of the care and services	1	1	0	0	0	0	1	3
Childhood Immunization Status	0	0	2	0	0	0	0	4
Adolescent Immunization Status	0	0	2	0	0	0	0	4
Prenatal Care - First Trimester	0	0	2	0	0	0	0	4
Well-Child Visits In The First 15 Months of Life (% with 6 or more)	0	0	2	0	0	0	0	4
Well-Child Visits In The Third, Fourth, Fifth And Sixth Year Of Life (% who received at least one well-child visit within the year)	0	0	2	0	0	0	0	4
Adolescent Well-Care Visit (% who received at least one comprehensive well-care visit within the year)	0	0	2	0	0	0	0	4
Breast Cancer Screening	0	0	1	1	0	0	1	4
Cervical Cancer Screening	0	0	1	1	0	0	1	4
Check-Ups After Delivery	0	0	1	1	0	0	1	4
Eye Exams For Diabetics	0	0	0	0	2	0	0	5
Follow-up After Hospitalization For Mental Illness within 30 days after discharge	0	0	0	2	0	0	0	5
Advising smokers to quit	0	0	1	1	0	0	1	5
Beta Blocker Treatment After a Heart Attack	0	0	0	1	1	0	1	5
32a. On how many of these visits were you advised to quit smoking by a doctor or other health professional in your plan?	0	0	1	1	0	0	1	5
Initiation of Prenatal Care within 6 weeks of enrollment (for enrollment between 6 and 44 weeks prior to live birth)	0	0	2	0	0	0	0	9
Board Certification (% - Primary Care, OB/GYN, Ped Specialists, Geriatricians, Other Physician Specialists)	0	2	0	0	0	0	0	9
Residency Completion (% - Primary Care, OB/GYN, Ped Specialists, Geriatricians, Other Physician Specialists)	0	2	0	0	0	0	0	9
Provider Compensation - Primary Care Physician Payment Mechanism	0	0	0	0	0	2	0	9
Arrangements with public health, educational, and social service organizations - tabular description	0	0	0	0	0	2	0	9
Enrollment by Payer	0	0	0	0	0	2	0	9
Total Enrollment	0	0	0	0	0	2	0	9
Rate Trends	0	0	0	0	0	2	0	9
High-Occurrence/High-Cost DRGs (discharges, days, total cost, discharges/1,000, avg cost/discharge, avg LOS)	0	0	0	0	0	2	0	9
Low Birth-Weight Babies (not required for reporting)	0	0	0	0	0	2	0	9
Years in Business/Total Membership	0	0	0	0	0	2	0	9

Table 4.1

Results of Expert Voting to Assign Measures to Categories

	ORIGINAL VOTES (FACCT and RAND) BY CATEGORY								FINAL CATEGORY ASSIGNMENT
MEASURE	Access and Service	Qualified Providers	Staying Healthy	Getting Better	Living with Illness	Exclude from Aggregation Strategy		Disagree? (1=yes)	1= NCQA Accred; 2= Access & Serv; 3 = Cons. Sat; 4 = Stay Healthy; 5 = GB/L w/ Ill; 9 = drop
Indicators of Financial Stability - Performance Indicators (total revenue, net income, net worth, debt-to-service ratio, administrative loss ratio, medical loss ratio, operating profit margin, overall profit margin)	0	0	0	0	0	2		0	9
Indicators of Financial Stability - Liquidity Indicators (days cash on hand, ration of cash to claims payable)	0	0	0	0	0	2		0	9
Indicators of Financial Stability - Efficiency Indicators (days in receivables, days in unpaid cash)	0	0	0	0	0	2		0	9
Indicators of Financial Stability -Statutory Indicators (state minimum reserve requirements, actual reserves held by plan)	0	0	0	0	0	2		0	9
Births and Average Length of Stay: Newborns	0	0	0	0	0	2		0	9
Mental Health Utilization - Inpatient Discharges and Average Length of Stay	0	0	0	0	0	2		0	9
Readmission for Selected Mental Health Disorders (within 90 and 365 days of discharge)	0	0	0	2	0	0		0	9
Chemical Dependency Utilization - Inpatient Discharges and Average Length of Stay	0	0	0	0	0	2		0	9
Readmission for Chemical Dependency (% readmitted w/in 90 and 365 days)	0	0	0	2	0	0		0	9
Outpatient Drug Utilization (Total cost and Average Cost per member by month and year)	0	0	0	0	0	2		0	9
Frequency Of Selected Procedures (Myringotomy, Tonsillectomy, Non-Obstetric Dilation and Curettage, Hysterectomy, Cholecytectomy, Laminectomy/Diskectomy, Angioplasty, Cardiac Catheterization, CABG, Prostatectomy)	0	0	0	0	0	2		0	9
Inpatient Utilization - General Hospital/Acute Care (discharges/1,000 member months, days, days/1,000 member months, ALOS) by Total inpatient, Medicine, Surgery, Maternity	0	0	0	0	0	2		0	9
Ambulatory Care (visits and visits/1,000 member months) by Total Outpatient Visits (excludes MH/CD), ER, Ambulatory Surgery/Procedures, Observation Room Stays	0	0	0	0	0	2		0	9
Inpatient Utilization - Nonacute Care (discharges, discharges/1,000 member months, days, days/1,000 member months, ALOS)	0	0	0	0	0	2		0	9
Discharge and Average Length of Stay - Maternity Care (total, vaginal, and cesarean deliveries)	0	0	0	0	0	2		0	9
1. Our records indicate that you are covered by [Health Plan Name]. Is this true?	0	0	0	0	0	2		0	9
2. How long have you been covered by [Health Plan Name]?	0	0	0	0	0	2		0	9
4. In the last 12 months, have you ever been a patient in a hospital overnight or longer?	#VALUE!	0	0	0	0	2		0	9
6. Why did you make visits that were NOT covered by your plan? Did not know the visit would not be covered Dissatisfied with plan doctors or care Plan did not cover the service Did not want to go through primary care doctor for a referral Could not get services quickly enough through plan The office or clinic is too far from home Did not want to take generic drugs Could not get a referral from plan Wanted a doctor not in the plan Service was covered by other insurance Some other reason Does not apply to me	0	0	0	0	0	2		0	9

Table 4.1
Results of Expert Voting to Assign Measures to Categories

	ORIGINAL VOTES (FACCT and RAND) BY CATEGORY								FINAL CATEGORY ASSIGNMENT
MEASURE	Access and Service	Qualified Providers	Staying Healthy	Getting Better	Living with Illness	Exclude from Aggregation Strategy		Disagree? (1=yes)	1= NCQA Accred; 2= Access & Serv; 3 = Cons. Sat; 4 = Stay Healthy; 5 = GB/L w/ Ill; 9 = drop
8. Thinking about your own health care and the services you received from [Health Plan Name] over the last 12 months, how would you rate the following? e. The cost you paid to belong to the plan	2	0	0	0	0	0		0	9
8. Thinking about your own health care and the services you received from [Health Plan Name] over the last 12 months, how would you rate the following? f. Amount you had to pay out-of-pocket (for example: co-payments, deductibles, payments for services not covered)	2	0	0	0	0	0		0	9
12. Over the last 12 months, did your plan's OVERALL performance get better, stay the same or get worse?	2	0	0	0	0	0		0	9
19.In general, would you say your health is: (excellent - poor)	0	0	0	0	0	2		0	9
20-25.(SF12) Normalized Health Functional Status	0	0	0	0	0	2		0	9
27a. Has a doctor EVER told you that you had any of the following conditions? Hypertension (sometimes called high blood pressure)	0	0	0	0	0	2		0	9
27b. Heart disease (like angina or heart failure)	0	0	0	0	0	2		0	9
27c. Diabetes (high blood sugar)	0	0	0	0	0	2		0	9
27d. Cancer (except skin cancer)	0	0	0	0	0	2		0	9
27e. Migraine (headaches)	0	0	0	0	0	2		0	9
28a. Do you NOW have any of the following conditions? Chronic allergies or sinus troubles	0	0	0	0	0	2		0	9
28b. Seasonal allergies or sinus troubles	0	0	0	0	0	2		0	9
28c. Arthritis or any kind of rheumatism	0	0	0	0	0	2		0	9
28d. Sciatica or chronic back problems	0	0	0	0	0	2		0	9
28e. Trouble seeing with one or both eyes, even when wearing glasses, blindness	0	0	0	0	0	2		0	9
28f. Chronic lung disease (like chronic bronchitis, asthma, or emphysema)	0	0	0	0	0	2		0	9
28g. Dermatitis or other chronic skin conditions	0	0	0	0	0	2		0	9
28h. Depression	0	0	0	0	0	2		0	9
28i. Ulcers in the stomach or duodenum, or heartburn	0	0	0	0	0	2		0	9
28j. Deafness or other trouble hearing with one or both ears	0	0	0	0	0	2		0	9
28k. Hemorrhoids	0	0	0	0	0	2		0	9
28l. Limitation in the use of an arm or leg (missing, paralyzed, or weakness)	0	0	0	0	0	2		0	9
29. Have you ever smoked at least 100 cigarettes in your entire life?	0	0	0	0	0	2		0	9
30. Do you now smoke every day, some days, or not at all?	0	0	0	0	0	2		0	9
31. How long has it been since you quit smoking cigarettes?	0	0	0	0	0	2		0	9
32. During the past 12 months, how many times have you visited a doctor or other health professional in your health plan?	0	0	0	0	0	2		0	9
URAC Accreditation	1	0	0	0	0	1		1	9
Adults' Access To Preventive/Ambulatory Health Services - % with visit during the reporting year	1	0	0	0	0	1		1	9
Children's Access to Primary Care Providers	1	0	0	0	0	1		1	9
Availability of Primary Care Providers	1	0	0	0	0	1		1	9
Availability of Behavioral Health Care Providers	1	0	0	0	0	1		1	9
Availability Of Obstetrical And Prenatal Care Providers	1	0	0	0	0	1		1	9
Low birth-weight deliveries at facilities for high-risk deliveries and neonates (not required for reporting)	0	1	0	0	0	1		1	9
Availability of Language Interpretation Services	1	0	0	0	0	1		1	9

Table 4.1
Results of Expert Voting to Assign Measures to Categories

	ORIGINAL VOTES (FACCT and RAND) BY CATEGORY								FINAL CATEGORY ASSIGNMENT
MEASURE	Access and Service	Qualified Providers	Staying Healthy	Getting Better	Living with Illness	Exclude from Aggregation Strategy		Disagree? (1=yes)	1= NCQA Accred; 2= Access & Serv; 3 = Cons. Sat; 4 = Stay Healthy; 5 = GB/L w/ Ill; 9 = drop
Disenrollment	1	0	0	0	0	1		1	9
Provider Turnover (Primary Care Physicians, Non-Physician Primary Care Providers)	1	0	0	0	0	1		1	9
Cesarean Section Rate	0	0	0	1	0	1		1	9
Vaginal Birth after Cesarean Section Rate (VBAC)	0	0	0	1	0	1		1	9
Mental Health Utilization - Percent of Members Receiving Inpatient, Day/Night Care, and Ambulatory Services	1	0	0	0	0	1		1	9
Chemical Dependency Utilization Percentage of Members Receiving Inpatient, Day/Night Care and Ambulatory Services	1	0	0	0	0	2		1	9
3. In the last 12 months, about how many visits did you make for yourself to a doctor's office, clinic or hospital emergency room? (This does not include staying overnight in a hospital or going to the dentist.)	1	0	0	0	0	1		1	9
5. In the last 12 months, did you make any visits for yourself to a doctor or other health care professional not covered by [Health Plan Name]?	1	0	0	0	0	1		1	9
26. Compared to one year ago, how would you rate your health in general now?	0	0	1	0	0	1		1	9

CHAPTER 5
AGGREGATION METHODS

A number of methodological issues arise when combining data to produce aggregate scores, including: handling missing data, standardization of scales, weighting, and incorporating uncertainty. Each of these issues is discussed in this section.

For each of these areas, a two-stage approach was taken. First, it was determined whether any of the various options for addressing an issue was likely to produce a different result. If the results were insensitive to the method of handling the result, then RAND recommended the method that was most defensible within the context of the project goals. If the results were sensitive to different methods, RAND presented the advantages and disadvantages of each option to the CARS Steering Committee for a policy decision. Frequently, choices required a compromise between statistical rigor, incentives for improved reporting in future years, and short run expediency.

HANDLING MISSING DATA

It was known that plans would be missing data in a variety of different patterns. Some plans might not report any items in a particular data source. Other plans might report some information but not all (e.g., some HEDIS measures might not be reported).

Six alternatives for handling missing data are possible: (a) do not report a score for the plan unless all items are available; (b) report the average of only those items that are reported; (c) impute the average value for the measure among plans that reported; (d) impute a score from regression analysis; (e) impute a zero value for the measure; (f) impute the lowest value reported by plans that provided data. Four of these alternatives (a, c, e, f) were modeled using 1997 HEDIS and MSS data and found that the method of handling missing data did affect the results. Each of the options is described followed by an analysis of the differences across the options.

Require Complete Reporting

Under this option, plans missing any individual measure in a scale would receive "NR" as their score. There was concern that this option would significantly decrease the number of plans for which a numeric summary score was available and that consumers might have difficulty comparing "not reporting" against an actual value. This strategy would also eliminate plans that had any data missing, thus potentially penalizing plans that provided most information.

Summarize Available Data

Another possibility is to summarize only the available data elements. For example, if the score is the average of 5 items and only 3 items are non-missing, the average of the non-missing items could be taken. The challenge with this approach is to make it "fair." Plans might be encouraged to report only the measures on which they perform well. If the items that a plan is missing are the most difficult to score well on, then averaging the easy items is unfair to the plans that report more completely. Fixing this problem quickly leads to elaborate imputation methods.

Impute the Average from Reporting Plans

This approach is frequently taken in research because it is simple and matches the mean of the imputed data to the mean of the raw data. The policy problem is that it does not encourage plans whose performance is below average to report because the imputed score they will receive is better than their actual score. Such incentives are inconsistent with the policy of the CARS Steering Committee which is to require all plans to report complete information.

Regression Imputation

More elaborate imputation methods are common in research analysis. The next level up on the complexity scale is regression imputation. Regression imputation uses the complete plans' data to develop a model of the relationship between the various items in a scale. This model is then used to impute missing values for the plans with incomplete data. This approach would have been further pursued if the sense of the committee hadn't already been headed in the direction of methods that strongly encouraged reporting. Regression imputation method also requires more lead time between receipt of the data and reporting than RAND anticipated having.

Imputing a Zero

In this option, plans not reporting a result for a measure would receive a value of zero for that measure. This option was viewed as both overly harsh, particularly to plans that report most of the data, and potentially not distinguishable from the "not reporting" option. It may, however, create incentives to report in future years.

Imputing the Lowest Observed Score

Under this option, plans not reporting a result for a measure would receive a value equal to the lowest number reported by any plan on that measure. This approach has the advantage of producing a nonzero score for a plan but placing performance at a low enough level that most plans should be able to improve their overall score by reporting

performance. This option was used at both the measure and data source level—that is, if a plan failed to report one or more measures within a category it would receive the lowest score reported for each measure. This was also true if a plan reported no measures within a category (e.g., if the plan had not fielded the Member Satisfaction Survey). For these items, the standard error was also set to zero, because we know the score with certainty.

Results

The two major options under consideration for handling missing data were mean and minimum value imputation. They also provide a sharp contrast in strategies. Mean imputation is an attempt, however simple, to fill in an estimate of what the missing value would have been had it been available. Minimum value imputation is not an attempt to estimate the missing value. Instead it is a policy choice designed to encourage more complete reporting in future years. Minimum value imputation sends the signal that reporting what you have is better than not reporting, no matter how poor your scores are.

Tables 5.1 through 5.5 present the results for the plans using both mean and minimum value imputation. The methods used to put plans into one of three result categories are discussed later in this chapter. Note that plans that land on the diagonal (i.e., 12, 90, 21 in Table 5.1) would receive the same number of stars under each method. Plans off the diagonal would receive different scores under the two methods. For example, in Table 5.1, of the 97 plans that would receive 2 stars under mean imputation 6 would receive 1 star and one would receive 3 stars under minimum value imputation. The 6 plans that moved down received lower scores because of missing data. The one plan that moved up did so because the minimum value imputation method moved the cut point lower.

In general, the choice of missing value methods has a substantial effect on plan results across all domains. Table 5.5 presents the correlations between the imputed scales. Note that the correlations are fairly high, 0.9 or better, with the exception of the Getting Better/Living With Illness category. This is a consequence of plans having more missing data in this category.

Table 5.1

Consumer Satisfaction: Comparison of Missing Value Imputation Methods

		Mean value imputation			
	Frequency	*	**	***	**Total**
Minimum value imputation	*	12	6	0	**18**
	**	4	90	5	**99**
	***	0	1	21	**22**
	Total	**16**	**97**	**26**	**139**

Table 5.2

Access and Service: Comparison of Missing Value Imputation Methods

		Mean value imputation			
	Frequency	*	**	***	**Total**
Minimum value imputation	*	20	8	0	**18**
	**	6	72	3	**99**
	***	0	3	27	**22**
	Total	**26**	**83**	**30**	**139**

Table 5.3

Staying Healthy: Comparison of Missing Value Imputation Methods

		Mean value imputation			
	Frequency	*	**	***	**Total**
Minimum value imputation	*	14	12	0	**26**
	**	10	79	1	**90**
	***	0	3	20	**23**
	Total	**24**	**94**	**21**	**139**

Table 5.4

Getting Better/Living with Illness: Comparison of Missing Value Imputation Methods

		Mean value imputation			
	Frequency	*	**	***	**Total**
Minimum value imputation	*	6	19	0	**25**
	**	9	77	2	**88**
	***	0	7	19	**26**
	Total	**15**	**103**	**21**	**139**

Table 5.5

Correlation of Scores for Minimum Value Imputation and Mean Value Imputation

Category	Correlation Coefficient
Consumer Satisfaction	0.928
Access and Service	0.913
Staying Healthy	0.906
Getting Better/Living with Illness	0.742

Policy Implications and Recommendations

The missing value imputation decision has the largest leverage of any technical decision in the CARS reporting process on the ultimate plan rankings. RAND recommended, and the CARS Steering Committee concurred, that a long run view was appropriate. Minimum value imputation was thought to provide the right incentives for reporting in future years.

STANDARDIZING SCALES

The individual measures that were being combined to create each of the summary categories were known, based on 1997 data, to have different means and variances. This potentially presents a problem for scaling in that it can make some measures have a greater (or lesser) effect on the results because of their distributional properties. RAND modeled the scales based on 1997 data and found few differences between the results based on standardized versus nonstandardized scales. However, because standardization is more defensible analytically, all items within scales were standardized.

Standardization is a simple calculation, but is frequently misunderstood due to its similarity to related statistical calculations. The idea is to transform item scores so that plans are ranked on a comparable scale across items. This prevents an item with a large range (say 1-100) from completely dominating an item with a small range (say 0-1).

Standardization is accomplished by dividing all of the plan values by the standard deviation of the plan values. For example, an item on the 0-100 scale might have a plan-to-plan standard deviation of 20. The standardized item would then run from 0-5. An item on the 0-1 scale could have a plan-to-plan standard deviation of 0.25. After standardization it would range from 0-4. The standardized scores are much more comparable than the original factor of 100 difference. Frequently the mean is subtracted prior to division by the standard error. This was not done in the analysis since only relative scores were used.

The plan-to-plan standard deviation is the standard deviation of the plan averages. Note that it is not the standard error of the mean or the standard deviation of the person level values. If the plan means were normally distributed most plans would fall between –2 and +2 after our standardization. One potential source of confusion is the superficial similarity of the standardizing calculation to the z-score test of difference from the mean. In the z-sore test, the divisor is a measure of sampling error rather than of distributional variability. Z-scores for plans could easily produce values far from the –2 to +2 range. Standardization is not meant to test the difference of the plan from the mean. It is just meant to put items on roughly comparable scales. The formula is shown below.

$$\textit{Standardized Item} = \frac{\textit{(Plan Value)}}{\textit{(Plan to Plan Standard Deviation)}}$$

The benefit of standardizing is that it makes it easier to compare items and to understand the meaning of weights applied to those items. By thinking in the standard deviation scale, simple rules of thumb based on the normal distribution make comparison easier. Thinking of a 1 standard deviation increase in each item is often easier than comparing a 35 point increase in a 100 point scale to a .012 increase in the mean of a dichotomous variable, for example.

Results

It was a surprise to see that standardization had little effect on plan rankings. Tables 5.6 – 5.9 show that, across the four domains, very few plans change the number of stars that they would receive ifstandardization is used or not. In retrospect, this probably should not have been a surprise. In most of the scales, the items were largely scored on the same scale prior to standardization. This similarity, combined with some correlation between items, mutes the effect of standardization. Note the very high (greater than 0.97) correlation between the standardized and unstandardized scales in Table 5.10.

Table 5.6

Consumer Satisfaction: Comparison of Standardization and No Standardization

		Without standardization			
	Frequency	*	**	***	**Total**
With standardization	*	18	0	0	**18**
	**	1	92	6	**99**
	***	0	0	22	**22**
	Total	**19**	**92**	**28**	**139**

Table 5.7

Staying Healthy: Comparison of Standardization and No Standardization

		Without standardization			
	Frequency	*	**	***	**Total**
With standardization	*	24	2	0	**26**
	**	5	82	3	**90**
	***	0	3	20	**23**
	Total	**29**	**87**	**23**	**139**

Table 5.8

Access and Service: Comparison of Standardization and No Standardization

		Without standardization			
	Frequency	*	**	***	**Total**
With standardization	*	23	5	0	**28**
	**	0	79	2	**81**
	***	0	0	30	**30**
	Total	**23**	**84**	**32**	**139**

Table 5.9

Getting Better/Living with Illness: Comparison of Standardization and No Standardization

		Without standardization			
	Frequency	*	**	***	**Total**
With standardization	*	24	1	0	**25**
	**	5	75	8	**88**
	***	0	1	25	**26**
	Total	**29**	**77**	**33**	**139**

Table 5.10

Correlation of Scores Between Standardization and No Standardization

Category	Correlation Coefficient
Consumer Satisfaction	0.991
Access and Service	0.988
Staying Healthy	0.971
Getting Better/Living with Illness	0.979

Policy Implications and Recommendations

Although there is very little leverage to the standardize vs. don't standardize decision, RAND recommends standardization. At the cost of a very small increase in complexity, standardization makes the meaning of weighting and the interpretation of scales more straightforward.

WEIGHTING

There were two major choices in constructing scales: (a) treat all measures as equally important thus allowing each measure to contribute an equal amount to the scale; (b) develop differential weights for measures within each scale. While there are a variety of different approaches that can be taken to developing weights, given the time constraints and experience of the RAND team with the measures, RAND proposed a set of expert weights. These weights are presented in Tables 5.11 through 5.14.

Table 5.11

Measures and Weights for Consumer Satisfaction Scale

Item (from Member Satisfaction Survey)	Weight
How much were you helped by the care you received?	1
How would you rate the overall quality of care and services?	1
All things considered, how satisfied are you with your current health plan?	1
Would you recommend your current health plan to your family or friends if they needed care?	1
Do you intend to switch to a different health plan when you next have an opportunity?	1

Table 5.12

Measures and Weights for Access and Service Scale

Scale/Item (from Member Satisfaction Survey)	Weight
Access	67%
How would you rate the ease of making appointments by phone?	1
How would you rate the length of time you had to wait between making an appointment for routine care and the day of your visit?	1
How would you rate the number of doctors you had to choose from?	1
Have you had a problem with delays in your medical care while you waited for approval by your health plan?	1
Have you had any difficulty in receiving care you and your doctor believed was necessary?	1
Have you ever not been able to get a referral to a specialist that you wanted to see?	1
How many days did you have to wait between making an appointment for routine care and the day you saw the provider?	1
How many days did you have to wait between making an appointment for a minor illness or injury and the day you saw the provider?	1
How many days did you have to wait between making an appointment for chronic or ongoing care and the day you saw the provider?	1
How many days did you have to wait between making an appointment for urgent care and the day you saw the provider?	1
How long did it usually take your provider's office to return a call?	1
How long did you have to wait to see your provider once you arrived at the office for an appointment?	1

Table 5.12

Measures and Weights for Access and Service Scale (cont.)

Scale/Item (from Member Satisfaction Survey)	Weight
Service	33%
How often did you see the same provider over the last 12 months?	1
How would you rate the thoroughness of the treatment you received?	1
How would you rate the attention given to what you had to say?	1
How would you rate the ease of choosing a personal physician?	1
How would you rate the amount of time you had with the doctor?	1
How would you rate the types of services covered?	1
How would you rate the availability of eligibility and coverage information?	1
How would you rate the availability of cost information?	1
How would you rate the time required to fill out paperwork?	1
Have you called with a complaint?	1
If you called with a complaint, how long did it take to resolve it?	1

The measures for Consumer Satisfaction each received equal weight. The measures shown in Tables 5.11 and 5.12 are not written exactly as presented in the survey instrument, but capture the basic content of the question. Each of the measures received equal weight within the Access and Service scale but the Access subscale was given twice as much weight as the Service subscale in producing the overall score for this dimension.

Table 5.13

Measures and Weights for the Staying Healthy Scale

Item	Weight
Childhood immunizations rate	15
Adolescent immunizations rate	5
Prenatal care initiated in the first trimester	14
Checkups after delivery	1
Six or more well child visits (0-15 months)	5
One or more well child visits (3-6 years)	5
One or more adolescent well care visits (12-21 years)	5
Breast cancer screening rate	30
Cervical cancer screening rate	20

Table 5.14

Measures and Weights for the Getting Better/Living with Illness Scale

Item	Weight
Follow-up after hospitalization for mental illness	20
Advising smokers to quit	30
Eye exams for persons with diabetes	20
Beta blocker treatment after heart attack	30

The measures in the Staying Healthy and Getting Better/Living with Illness scales were weighted to reflect the expected health benefits to the population of improved performance on the measure. This information was derived from a previous review of the scientific literature conducted by RAND[4].

Results

The 1997 Quality Compass data were used to illustrate the effect of weights on results. There is a small effect of weighting on plan scores. Tables 5.15 – 5.18 compare the results for plans with and without weighting. The Staying Healthy/Living with Illness scale shows some differences as a result of weighting. Note in Table 5.19 that all of the correlations (weighted vs. unweighted) are above 0.92.

---

[4] McGlynn, Keeler, Tseng, et al. *Health Benefits and Cost Effectiveness of Attaining the Goals set By the HEDIS 2.5 Quality Indicators*, unpublished.

Table 5.15

Consumer Satisfaction: Comparison of Weights versus No Weights

		No Weights			
	Frequency	*	**	***	**Total**
Weights	*	18	0	0	**18**
	**	0	99	0	**99**
	***	0	0	22	**22**
	Total	**18**	**99**	**22**	**139**

Table 5.16

Access and Service: Comparison of Weights versus No Weights

		No Weights			
	Frequency	*	**	***	**Total**
Weights	*	26	2	0	**28**
	**	0	79	2	**81**
	***	0	3	27	**30**
	Total	**26**	**84**	**29**	**139**

Table 5.17

Staying Healthy: Comparison of Weights versus No Weights

		No Weights			
	Frequency	*	**	***	**Total**
Weights	*	22	4	0	**0**
	**	4	81	5	**5**
	***	0	5	18	**18**
	Total	**26**	**90**	**23**	**139**

Table 5.18

Getting Better/Living with Illness: Comparison of Weights versus No Weights

		No Weights			
	Frequency	*	**	***	**Total**
Weights	*	25	0	0	**25**
	**	3	85	0	**88**
	***	0	0	26	**26**
	Total	**28**	**85**	**26**	**139**

Table 5.19

Correlation of Scores with Weights vs. Without Weights

Category	Correlation Coefficient
Consumer Satisfaction	1.00
Access and Service	0.998
Staying Healthy	0.927
Getting Better/Living with Illness	0.992

Policy Implications and Recommendations

Weighting items does have a small effect on the results. The expert weights were reviewed by the CARS Steering Committee and approved with some modifications. The final measures in each scale and associated weights are shown in Tables 5.11 – 5.14. Weighting does focus attention on the importance of the processes measured in each scale.

STANDARD ERROR OF THE SCALE SCORES

Measurement of health plan performance is an imperfect science; a variety of errors occur in measuring performance. In order to provide valid information for users of performance measures, we must find methods to incorporate some of the uncertainties about the results into the reporting.

Item level measurement standard errors

We had to calculate standard errors from the available confidence intervals for items when the standard errors were not reported by plans. This was accomplished by solving the HEDIS formulas for the sample size and calculating the standard errors. For a few plans, we obtained rates of performance but not standard errors or sample sizes. The required sample sizes for these plans was imported by averaging the minimum of the plans' reported N or 411, whichever was smaller. This was the best available estimate of the hybrid method N. The smaller N was chosen to avoid assigning a small standard error when we were uncertain of the true sample size.

Scale standard errors

The scales were ultimately the weighted sum of the individual items. The weights and the item level standard errors were used to calculate an overall standard error for the scale.

Standard error information is typically used in one of two related ways. First, standard errors can be used to produce confidence intervals around the scale scores. This would be most useful when presenting numeric scale scores or graphs that represent these scores. Second,

standard errors can be used to perform statistical tests. This would be most useful when the presentation was making some comparative claim like: "performance is above the mean." There is a strong relationship between the two approaches. In the statistical star method presented below we will actually perform a test by constructing a confidence interval and seeing if that interval touches a cutpoint.

CUTPOINTS

The 1997 Quality Compass data were used to model different choices of cutpoints, comparing performance in the top one-third versus the top one-quarter.

Within each scale, the 25th and 75th percentiles were calculated across the distribution of scores for all plans contracting with one or more CARS Steering Committee members. This exercise was then repeated this using the 33rd and 67th percentiles. These established the comparison points that plans were tested against to determine their star rating in each category.

A statistical star system was used to incorporate uncertainty by using the standard errors of the scores. The method using the 25/75 cutpoints is described here. The 33/67 scale was produced in the same way using the different cutpoints. If a plan's score fell below the 25th percentile nationally, then a 95% confidence interval for the plan was constructed. If this confidence interval was entirely below the 25th percentile, the plan was given one star. If a plan's score was above the 75th percentile nationally, then the plan's 95% confidence interval was also constructed. If this confidence interval was entirely above the 75th percentile, the plan received three stars. All other plans received two stars. The logic was to give one- or three-star scores to only those plans where we were fairly certain that the plans were in the bottom or top quarters of the national plan distribution.

Results

Tables 5.20 – 5.23 show the relationships between the star assignments using the two different cutpoints. As you would expect, the 33/66 method assigns more plans to the one and three star categories.

Table 5.20

Consumer Satisfaction: Cutpoint Comparisons

		33/67 Cutpoints			
	Frequency	*	**	***	**Total**
25/75	*	18	0	0	**18**
Cutpoints	**	10	76	13	**99**
	***	0	0	22	**22**
	Total	**28**	**76**	**35**	**139**

Table 5.21

Staying Healthy: Cutpoint Comparisons

		33/67 Cutpoints			
	Frequency	*	**	***	**Total**
25/75	*	26	0	0	**26**
Cutpoints	**	10	74	6	**90**
	***	0	0	23	**23**
	Total	**36**	**74**	**29**	**139**

Table 5.22

Access and Service: Cutpoint Comparisons

		33/67 Cutpoints			
	Frequency	*	**	***	**Total**
25/75	*	28	0	0	**28**
Cutpoints	**	8	60	13	**81**
	***	0	0	30	**30**
	Total	**36**	**60**	**43**	**139**

Table 5.23

Getting Better/Living with Illness: Cutpoint Comparisons

		33/67 Cutpoints			
	Frequency	*	**	***	**Total**
25/75	*	25	0	0	**25**
Cutpoints	**	12	67	9	**88**
	***	0	0	26	**26**
	Total	**37**	**67**	**35**	**139**

Policy Implications and Recommendations

The CARS Steering Committee elected to characterize performance using symbols (rather than numbers or graphs) and chose stars as the symbol. Ford and DaimlerChrysler had previously used stars to characterize plan performance; GM had used triangles. Plans were assigned one, two, or three stars within each reporting category based on a statistical test (we referred to this approach as "statistical stars"). By contrast, the Ford 1997 system assigned stars without a statistical test; the plan distributions were divided into three equal sized groups. The statistical test provides a more defensible method of assigning plans to result groups because it takes account of the uncertainty that results from drawing samples to estimate performance. Because the CARS Steering Committee wanted the stars to clearly distinguish among different levels of performance the more stringent cutpoints were selected (25th and 75th percentiles).

CHAPTER 6
REPORTING STRATEGY

The Steering Committee also had to agree on how the results would be communicated both internally to benefits managers and externally to employees during open enrollment. The Steering Committee agreed that the presentation of results should be uniform across the open enrollment materials for each of the three automobile manufacturers. This included the order in which categories were presented and the method of communicating scores (i.e., stars).

Two major options were considered: (a) presenting numeric scores or (b) representing results with symbols. The CARS Steering Committee elected to use symbols and chose stars as the symbol. The major reason for this choice was that all three automobile companies had previously used symbols in reporting performance results. There is some evidence that consistency in reporting improves the likelihood that people will use the information. Two of the companies (Chrysler and Ford) had used stars previously; GM had used triangles. All three companies had used three levels of performance to report results. None of the companies had used a statistical method for determining the placement of plans in result groups.

The 1997 Quality Compass data were used to model different choices of cutpoints, comparing performance in the top one-third versus the top one-quarter. Because the CARS Steering Committee wanted the stars to clearly distinguish among different levels of performance the more stringent cutpoints were selected (25^{th} and 75^{th} percentiles).

Within each scale, the 25^{th} and 75^{th} percentiles were calculated across the distribution of scores for all plans contracting with one or more CARS Steering Committee members. If a plan's score fell below the 25^{th} percentile nationally, we constructed a 95% confidence interval for the plan. If this confidence interval was entirely below the 25^{th} percentile, the plan received one star. If a plan's score was above the 75^{th} percentile nationally we also constructed a 95% confidence interval for the plan. If this confidence interval was entirely above the 75^{th} percentile, the plan received three stars. All other plans received two stars. The logic was only to give one or three star scores to plans where we were fairly certain that the plans were in the bottom or top quarters of the national plan distribution. This method was referred to as "statistical stars" to reflect the use of statistical principles to make assignments to result categories.

The Steering Committee valued the opportunity to incorporate statistical uncertainty into the performance reports—the statistical star model facilitated this objective.

The primary media for transmitting results were printed brochures that were part of the open enrollment materials. Employees received information on those health plans in their geographic area which were open for enrollment. In most areas, the performance reports were not available for the indemnity or traditional plans. GM did provide survey data for scoring traditional indemnity plans and preferred provider organizations in three states (Michigan, Ohio, Indiana) on two of the performance categories (Consumer Satisfaction and Access and Service).

In addition to open enrollment materials, information was made available to employees on internal Web sites. The Greater Detroit Area Health Council also reported the results on its Web page for plans operating in Southeast Michigan (http://www.hiag.org/qat/hmoguide.htm).

A sample report card is shown in Figure 6.1.

Sample CARS '98 Report

Health Plan	NCQA Accreditation Status	Consumer Satisfaction	Access and Service	Staying Healthy	Getting Better/Living with Illness
HMO A	Accredited	★★	★★	★★	★★
HMO B	Accredited	★★	★★★	★★	★★
HMO C	Accredited	★★	★★	★★★	★
HMO D	Accredited	★★★	★★	★★	★★★
HMO E	Accredited	★★	★★★	★★★	★★
HMO F	Not Accredited	★	★	★★★	★

★★★ Significantly above average
★★ Average
★ Significantly below average

Figure 6.1—Sample Report Card

CHAPTER 7
OBTAINING DATA FROM HEALTH PLANS

The three data sources for the CARS model are HEDIS 3.0/1998, the NCQA Membership Satisfaction Survey (MSS), and NCQA Accreditation Status. The first two data sources are available from NCQA's Quality Compass. NCQA's Quality Compass is a national CD-ROM database of plan specific performance information including nearly 300 health plans. However, in consultation with NCQA, RAND and the Steering Committee concluded that a direct data feed from NCQA would not be available in time to produce scores on the schedule required by GM and DaimlerChrysler for their open enrollment materials. Further, NCQA would only be able to provide data for the health plans that reported to Quality Compass for public release.[4] Of the 131 unique managed care plans with which DaimlerChrysler, Ford and GM contract, 76 plans publicly reported their HEDIS 3.0/1998 performance to Quality Compass. Due to the time constraints as well as the fact that not all plans report to Quality Compass for public release, the data were obtained directly from the health plans.

PROCESS FOR OBTAINING DATA

A combined list of plans that contracted with one or more of the companies was developed to coordinate the data requests. Table 7.1 shows the number of plans each company contracted with and the amount of overlap in the contracts. All plans were contacted in writing during May 1998 and informed that June 15, 1998 was the deadline to submit HEDIS and MSS data. GM took the lead on requesting data from the health plans with which it held contracts, many of which also held contracts with Ford and/or DaimlerChrysler (see Table 7.1). Representatives from Ford and DaimlerChrysler contacted the remaining health plans.

---

[4] Health plans have the option to report their performance to NCQA for use in calculating national average performance without allowing their data to be available for public release to consumers, purchasers, or researchers.

Table 7.1

Number of Health Plan Contracts by Company and Degree of Overlap

Contract Description	NUMBER OF PLANS*
Chrysler (all plans)	24
Ford (all plans)	73
GM (all plans)	108
Plans that contract with all three	14
Plans that contract with two	46

* Includes some plans that were being considered for inclusion for the first time; in the case of multiple contracts with a plan, the plan is only counted once.

NCQA allowed the plans that intended to report to Quality Compass to send RAND copies of the HEDIS 3.0/1998 Data Submission Tool (DST) for use in the CARS project. The DST is an electronic file and was submitted by the plans to RAND either on disk or via electronic mail. Plans not reporting to Quality Compass could not legally use the DST, so a separate survey for the CARS project was developed to obtain the relevant HEDIS and MSS results. The CARS data collection survey was available to the health plans in both an electronic and paper form. Representatives from the health plans were able to submit the completed survey via fax or electronic mail.

The MSS data were not included in the DST. The standard format in which the MSS data were received was a text file that met NCQA specifications to report plan-level MSS results. If plans did not have the MSS text file available for submission, they were able to use the CARS data collection survey. Table 7.2 shows the number of plans that submitted HEDIS and MSS data in the various formats.

Table 7.2

Number of Health Plans by Submission Format

	Number of Plans by Type of Data	
Data Submission Format	**HEDIS**	**MSS**
Preferred format (e.g., DST and text file)	80	43
CARS survey	33	49
Member level data	--	3
Hard-copy	15	31
No data submitted	3	5
TOTAL	131	131

HEDIS data were received for 128 health plans. Of the 48 health plans that did not submit HEDIS data via the DST, 33 plans used the CARS survey. The remaining plans submitted a hard-copy summary of their HEDIS performance results. A total of 126 plans submitted MSS data. Of the 83 plans that did not submit the standard text file, 49 plans used the CARS data collection survey. Three plans submitted member level data files that required RAND to generate the summary statistics used in the aggregation method. The remaining plans submitted a hard-copy summary of their MSS results.

NCQA accreditation status is available on NCQA's Web page (www.ncqa.org). The Accreditation Status List (ASL) is updated by NCQA on a monthly basis. Consequently, there is a potential lag between a change in accreditation status and when that information is available through NCQA's Web page. The results used by the auto manufacturers in their open enrollment materials were based on information on the Web as of August 3, 1998.

DATA VERIFICATION

The initial requests for data were sent to the health plans to coincide with the date that initial Quality Compass reports were due to NCQA. Because NCQA engages in a process of reviewing results and correcting errors in the initial data submissions, RAND chose to mirror this process. RAND developed a form to feedback the raw inputs used in the development of the performance scores (Appendix A). The verification reports were sent to all plans submitting data to ensure that the information was correct. In addition, five days before the deadline for data submission all plans received an information sheet explaining the consequences of not reporting (i.e., imputation of the lowest reported value for missing data). Missing data were highlighted and the plans had a final opportunity to submit data. Table 7.3 shows the data collection timeline.

Table 7.3
Data Collection Timeline

Date	Milestone	Data Received (% of plans) HEDIS	MSS
5/22/98	Data request memo to plans	--	--
6/15/98	Submissions due to NCQA and RAND	0	0
7/98 – 8/98	RAND sends data verification reports	--	--
7/17/98	Corrected submissions due to NCQA and RAND	43	37
7/27/98	RAND memo to plans stating final deadline of August 1st	62	55
8/03/98	FINAL deadline for data submissions	92	79
8/14/98	Cutpoints determined	95	94

SUMMARY OF DATA COMPLETENESS

Twelve HEDIS measures were included in the CARS model. For those plans that submitted HEDIS data, there was an average of one and a maximum of seven missing measures. The number of missing measures did vary by data submission format. Plans that used the DST were missing less than one measure on average, while plans that used alternative submission formats were missing 1.8 measures on average. Table 7.4 shows the completeness of reporting for each HEDIS measure.

Table 7.4

Percent of Plans Reporting HEDIS Measures by Submission Type

Measure	Percent of Plans Reporting a HEDIS Measure by Submission Type		
	DST	Other submission format	Total
Childhood immunization rate	93	94	93
Adolescent immunization rate	84	65	77
Prenatal care – first trimester	98	100	99
Check-ups after delivery	96	96	96
Well-child visits in first 15 months of life	96	75	89
Well-child visits in 3-6th year of life	96	90	94
Adolescent well-care visits	95	92	94
Breast cancer screening	100	100	100
Cervical cancer screening	100	100	100
Eye exams for diabetics	96	96	96
Follow-up after mental health hospitalization	95	56	81
Beta blocker treatment after a heart attack	81	58	73

The CARS model included 29 items from the MSS. On average, plans were missing less than one MSS measure. The maximum number of missing MSS measures was 27; this came from a plan that submitted the results of only two global assessment questions. The plan either did not field the entire MSS or failed to submit the results of the entire survey. None of the plans that used the standard text file data submission format were missing any of the MSS items. One of the performance measures, advice to quit smoking, is derived from the MSS.

CHAPTER 8
RESULTS

The three auto manufacturers spend approximately $8 billion annually to provide health insurance for over 2 million people, and contract with 131 managed care in 34 different states. Thus, the potential effect of this national-level effort to assess health plan performance is quite large. In several Midwestern states, these companies represent among the largest employers operating and thus are a significant portion of the business for health plans delivering services in these markets.

Results of the health plan performance evaluation were published in 1999 open enrollment materials for the employees of DaimlerChrysler, Ford, and GM. The results for Southeastern Michigan were also made available on GDAHC's Web page (http://www.hiag.com).

Table 8.1 shows the distribution of stars across the plans by performance category. For each reporting category, at least one-half of the health plans received two stars. There was a range of 16 to 21 percent of the plans performing below the 25th percentile for each of the reporting categories; these plans received one star. For each reporting category, 14 to 20 percent of the plans received three stars because they clearly performed above the 75th percentile.

Table 8.1
Distribution of Plan Results by Performance Category

Category	Percent of plans with one star	Percent of plans with two stars	Percent of plans with three stars
Consumer Satisfaction	16	70	14
Access and Service	22	58	20
Staying Healthy	16	69	15
Getting Better/Living with Illness	21	59	20

Due to the minimum imputation rule for missing values, a plan's performance within a category was diminished if component items were not reported. Table 8.2 shows the average number of missing items for each performance category by the number of stars that the plan received. There was one plan that received three stars in the Staying Healthy/Living with Illness category even though it was missing one component of the

reporting category; no other plans with missing data received three stars in the category where the data were missing.

Table 8.2

Average Number of Missing Items by Score in each Performance Category

Category	Average Number of Missing Items by Score			Total Number of Items in Category
	One star	Two stars	Three stars	
Consumer Satisfaction	1.67	0	0	5
Access and Service	5.9	0.1	0	23
Staying Healthy	3.9	0.5	0.1	9
Getting Better/Living with Illness	2.3	0.5	0	4

Plans showed differential performance across the four reporting categories. Table 8.3 shows the percent of plans that received the same score in each category as well as the percent of plans that received a one star or three star score in at least one category. Five percent of the plans received one star in at least one category and three stars in another category. Four percent of the plans received one star in every category, 25 percent of the plans received two stars in all four reporting categories, and three percent of the plans received three stars in every category.

Table 8.3

Percent of Plans by Score

Score	Percent of plans
One star in every category	3.8
Two stars in every category	25.2
Three stars in every category	3.1
One star in at least one category	42.1
Three stars in at least one category	36.9

Table 8.4 gives the results of CARS '98 by health plan name. NCQA Accreditation is shown as 1 for accredited and 0 for not accredited. While NCQA has many more categories of accreditation, the group elected this dichotomous approach to simplify reporting. The next four columns give the actual score for each plan by category. These scores reflect the performance result times the weight and are not directly interpretable. The last four columns give the number of stars that were assigned

based on the score, the confidence interval around that score and its relationship to the selected cutpoints.

Table 8.4
Coordinated Autos/UAW Reporting System
(CARS '98) FINAL RESULTS

Health Plan Name	NCQA Accreditation	Consumer Satisfaction Stars	Access & Service Stars	Staying Healthy Stars	Getting Better/Living with Illness Stars	Consumer Satisfaction Percentile	Access & Service Percentile	Staying Healthy Percentile	Getting Better/Living with Illness Percentile
Aetna U.S. Healthcare - Dallas/Fort Worth Metroplex	1	1	1	2	1	9	11	34	17
Aetna U.S. Healthcare - Eastern/Central Pennsylvania	1	2	3	2	2	65	83	32	63
Aetna U.S. Healthcare - Greater Houston Area	1	1	1	1	1	11	12	16	22
Aetna U.S. Healthcare - New York	1	2	2	1	2	59	51	13	54
Aetna U.S. Healthcare - Northern New Jersey	1	2	2	1	2	41	62	12	65
Aetna U.S. Healthcare - Southern New Jersey/Delaware	1	2	3	2	2	84	93	23	63
Anthem BCBS of OH (Health Maintenance Plan)	1	2	3	2	2	79	79	58	60
Anthem PMP (Ohio PPO)	0	2	2	1	1	39	47	3	3
Anthem PPN (IN PPO)	0	2	2	1	1	56	67	3	3
AvMed Health Plan - Florida	1	2	2	2	2	72	45	45	54
BCBS Michigan - PPO	0	2	2	1	1	47	60	3	3
BCBS Michigan - Traditional	0	2	3	1	1	47	81	3	3
Blue Care Health Plan	1	2	3	3	2	81	88	88	80
Blue Care Network, East Michigan Region	1	2	2	2	3	45	64	84	98
Blue Care Network, Mid Michigan Region	1	2	3	3	2	69	79	86	56
Blue Care Network, Southeast Michigan Region	1	2	2	2	2	25	46	28	74
Blue Care Network, West Michigan Region	1	2	2	2	2	61	75	83	73
Blue Choice (Arizona)	0	1	1	1	1	3	2	20	13
Blue Cross Blue Shield of Kansas City	1	2	2	2	2	31	50	27	42
Blue Plus (Minnesota)	0	2	2	2	2	58	74	56	72
BlueChoice Healthcare Plan	1	2	2	2	2	68	61	25	47
BlueLincs HMO	1	2	2	2	2	30	43	36	39
CIGNA HealthCare of Arizona - Tucson	1	2	2	2	2	46	28	61	35
CIGNA HealthCare of Florida	1	1	1	2	2	8	15	53	61
CIGNA HealthCare of Georgia	1	1	1	1	1	9	15	15	15
CIGNA HealthCare of Louisiana	0	1	1	2	1	6	9	24	15
CIGNA HealthCare of North Texas - Dallas	1	2	2	2	1	48	49	43	24
CIGNA HealthCare of OH/Cincinnati	1	2	2	2	1	36	59	40	18
CIGNA HealthCare of Ohio - Columbus	1	1	2	2	1	11	25	44	20
CIGNA HealthCare of South Texas - Houston	1	2	2	2	1	57	52	26	14
CIGNA HealthCare of Tennessee	1	2	2	2	1	43	40	36	16
Capital District Physicians Health Plan	1	2	3	3	2	77	91	87	77
Care Choices HMO/Mercy Health Plans	1	2	2	2	2	73	68	68	30
Care Choices of Iowa	1	2	2	2	2	73	68	68	30
Community Health Plan	1	2	2	3	3	64	69	93	92
Compcare - Wisconsin	0	1	1	2	3	6	7	54	87
ConnectiCare, Inc.	1	3	3	3	3	95	78	91	85
Dean Health Plan, Inc.	1	3	3	2	2	85	86	72	75
Exclusive Healthcare, Inc. - Omaha	1	2	2	2	2	52	61	56	68
Fallon Community Health Plan	1	3	3	3	3	93	94	99	99
Family Health Plan Cooperative	0	2	2	2	2	24	47	77	69
Finger Lakes Health Insurance Co.	1	2	3	3	3	77	84	96	95
Freestate HP/CareFirst (includes Columbia Medical Plan)	1	2	2	2	2	53	42	54	57

Table 8.4
Coordinated Autos/UAW Reporting System
(CARS '98) FINAL RESULTS

Health Plan Name	NCQA Accreditation	Consumer Satisfaction Stars	Access & Service Stars	Staying Healthy Stars	Getting Better/Living with Illness Stars	Consumer Satisfaction Percentile	Access & Service Percentile	Staying Healthy Percentile	Getting Better/Living with Illness Percentile
Grand Valley Health Plan, Inc.	0	2	3	3	3	54	81	88	96
Group Health Cooperative of Puget Sound	1	2	2	2	3	34	54	74	97
Group Health Cooperative of South Central Wisconsin	1	3	3	3	2	94	89	97	64
Group Health Northwest	1	2	2	2	2	74	73	81	78
Group Health Plan, Inc. - St. Louis	1	2	2	2	2	59	52	71	53
HEALTHPARTNERS	1	2	2	2	2	63	65	67	70
HIP Health Plan of Florida, Inc.	1	2	2	2	2	49	22	33	49
HIP Health Plan of New Jersey	1	2	1	2	2	26	21	22	52
HIP Health Plan of New York	1	2	1	2	2	18	13	34	58
HMO Blue (BCBS Mass)	1	3	2	3	3	99	36	95	93
HMO Blue Nashville	0	1	1	1	1	3	2	3	3
HMO Blue of New Jersey	1	2	1	1	2	22	18	20	34
HMO Choice	1	3	3	2	2	90	97	73	72
HMO Colorado	0	2	2	1	2	45	32	13	52
HMO El Paso	0	2	1	1	1	28	10	19	8
HMO Health Ohio	0	2	2	1	2	51	45	9	34
Harris Methodist Health Plan	1	2	2	2	3	38	27	55	82
Harvard Pilgrim Health Care	1	2	2	3	3	76	72	98	89
Health Alliance Medical Plans, Inc.	1	3	3	3	3	91	93	90	83
Health Alliance Plan of Michigan	1	2	2	2	3	60	50	65	81
Health Care Plan, Inc. - NY	1	2	3	3	3	43	90	89	95
Health Net - CA	1	2	2	1	2	50	31	10	36
Health Options - Central FL / BCBS	1	1	1	2	2	14	16	38	46
Health Options - Northern FL / BCBS	1	2	2	2	2	81	70	57	56
Health Options - Southern FL / BCBS	1	2	1	2	2	35	20	25	45
Health Options of Fla. / Capitol Health	1	3	3	1	1	97	86	3	20
HealthAmerica Eastern PA	1	1	1	2	2	3	4	76	79
HealthAmerica Pittsburgh	1	1	1	3	3	3	4	97	88
HealthPlus of Michigan	1	3	3	3	2	93	77	86	79
HealthSource of Indiana	0	1	2	2	2	13	39	31	29
HealthSource of Kentucky	0	3	3	2	1	92	88	39	22
Humana Chicago	1	1	1	2	2	10	13	29	70
Humana Choice Care	1	3	3	2	2	84	87	75	38
Humana Daytona	1	2	2	2	1	19	35	77	25
Humana Fort Walton	1	2	2	2	2	21	23	30	32
Humana Jacksonville	1	2	2	2	1	31	24	38	21
Humana Kansas City	1	1	2	2	3	13	34	80	90
Humana Louisville	0	2	2	2	2	40	56	31	55
Humana Milwaukee	0	2	2	2	2	34	58	63	31
Humana Orland / Central Fla.	1	2	2	2	2	19	35	52	28
Humana PCA Texas	1	2	2	1	1	37	33	7	10
Humana San Antonio	1	2	1	2	2	25	18	49	40
Humana South Florida	1	1	1	2	2	7	17	50	67

Table 8.4
Coordinated Autos/UAW Reporting System
(CARS '98) FINAL RESULTS

Health Plan Name	NCQA Accreditation	Consumer Satisfaction Stars	Access & Service Stars	Staying Healthy Stars	Getting Better/Living with Illness Stars	Consumer Satisfaction Percentile	Access & Service Percentile	Staying Healthy Percentile	Getting Better/Living with Illness Percentile
Humana Tampa	1	2	1	2	2	16	14	70	26
Indiana Anthem BCBS (Traditional)	0	2	2	1	1	38	63	3	3
Intergroup of Arizona	1	1	1	2	2	12	11	66	71
Kaiser Foundation Health Plan of Colorado	1	2	2	3	3	63	63	85	88
Kaiser Foundation Health Plan of Georgia, Inc.	1	2	2	2	2	61	57	65	75
Kaiser Foundation Health Plan of Kansas City, Inc.	1	2	2	2	3	33	37	51	86
Kaiser Foundation Health Plan of North Carolina	1	2	2	2	3	66	48	62	81
Kaiser Foundation Health Plan of Ohio	1	2	2	2	3	32	38	81	93
Kaiser Foundation Health Plan of So. California	1	2	2	3	3	70	41	92	84
Kaiser Foundation Health Plan of Texas	1	2	2	2	3	55	76	45	86
Kaiser Foundation Health Plan of the Mid-Atlantic States	1	2	2	3	2	27	29	90	77
Kaiser Foundation Health Plan of the Northwest	1	2	2	2	3	36	44	61	90
Kaiser Foundation Health Plan, Inc. Hawaii Region	1	3	3	2	3	95	92	72	97
Kaiser Foundation Health Plan, Inc. No. California	1	2	2	2	3	29	30	70	84
Kaiser Foundation Health Plan, Northeast Region	1	2	2	3	3	64	69	93	92
Lovelace	1	2	2	2	2	71	53	47	65
M Plan	1	2	2	2	2	80	75	79	62
M-Care	1	3	2	3	2	89	72	94	38
Maxicare Indiana, Inc.	0	2	2	1	2	68	66	11	37
Medical Value Plan	1	3	3	2	2	98	99	41	40
Mercycare Health Plans	0	2	3	2	2	70	85	60	50
NYLCare of the Mid-Atlantic	1	1	1	2	1	3	6	48	6
Ochsner Health Plan	0	3	2	2	2	86	65	43	59
Ohio Anthem BCBS (Traditional)	0	2	2	1	1	15	40	3	3
Omnicare	1	2	1	1	2	17	20	8	33
Optima Health Plan	1	2	2	2	1	79	71	52	12
PARTNERS National Health Plans of Indiana	1	1	1	2	1	3	2	29	7
PHP - NY	0	3	3	2	2	83	77	75	27
PacifiCare of California	1	2	2	2	2	54	26	40	50
PacifiCare of Nevada	1	2	2	1	2	67	43	14	59
PacifiCare of Oklahoma	1	2	1	2	2	75	8	47	48
PacifiCare of Oregon	1	2	1	2	2	50	9	50	76
PacifiCare of Texas	0	2	1	1	1	29	6	11	9
Paramount Health Care	1	3	3	2	2	87	90	63	47
Physicians Health Plan of Mid Michigan, Inc.	0	3	3	2	2	96	97	79	36
Physicians Health Plan of South Michigan, Inc.	0	3	3	1	2	88	96	18	41
Physicians Health Plan of Southwest MI, Inc.	0	3	3	1	2	86	82	18	45
Physicians Health Plan of West Michigan, Inc.	0	3	3	2	2	90	95	46	43
Preferred Care, Inc.	1	3	3	3	3	97	98	95	94
PrimeCare Select	1	3	3	2	2	88	80	69	61
Priority Health - Michigan	1	2	3	2	2	82	84	82	51
Prudential HealthCare - Florida (Jacksonville)	1	2	2	2	2	75	56	78	44
Prudential HealthCare - Missouri (St. Louis)	1	2	2	2	2	62	55	42	25

Table 8.4
Coordinated Autos/UAW Reporting System
(CARS '98) FINAL RESULTS

Health Plan Name	NCQA Accreditation	Consumer Satisfaction Stars	Access & Service Stars	Staying Healthy Stars	Getting Better/Living with Illness Stars	Consumer Satisfaction Percentile	Access & Service Percentile	Staying Healthy Percentile	Getting Better/Living with Illness Percentile
Prudential HealthCare - Oklahoma	1	2	2	2	2	52	54	84	27
Prudential HealthCare - Tennessee (Memphis)	1	2	2	1	1	44	31	17	11
Prudential HealthCare - Tennessee (Nashville)	1	2	2	2	1	27	38	21	18
Prudential HealthCare - Texas (Austin/Central Texas)	1	2	2	2	2	40	36	35	43
Prudential HealthCare - Texas (Houston)	1	2	1	2	2	23	19	59	66
Prudential HealthCare - Texas (San Antonio)	1	2	2	2	2	56	59	22	23
Prudential HealthCare HMO - Georgia	1	2	2	1	1	20	27	15	13
QualMed Plans for Health, Inc (PA)	1	1	2	1	1	15	29	9	11
Rockford	0	1	1	1	1	3	2	3	3
Rush Prudential HMO	1	1	1	2	1	3	2	37	9
SelectCare HMO	1	2	2	2	2	18	22	64	68
The HMOs of Blue Cross and Blue Shield of Illinois	0	2	2	2	2	42	25	27	29
Total Health Care (Detroit)	0	2	1	1	1	22	5	3	3
United HealthCare of Upstate New York	0	2	3	2	1	78	95	59	19

CHAPTER 9
DISCUSSION AND NEXT STEPS

The purpose of this report is to document in detail the process, considerations, and decisions that were made in developing a coordinated strategy for reporting on health plan performance. The intent of the CARS Steering Committee from the beginning was to place these methods in the public domain so that they would be accessible to others interested in pursuing a similar strategy for reporting.

IMPORTANCE OF A COORDINATED STRATEGY

Just a decade ago, no systematic information was routinely available on the quality of health services being delivered across the country. With the introduction of HEDIS, this has begun to change. We have rapidly moved from not having enough information to having too much to make use of in decision making. Thus, strategies for summarizing results have become of great interest. The lay press in particular have undertaken efforts to provide simple summary scores on health plan performance using available information. Unfortunately, differences in the methods by which summary scores are derived can lead to differences in the conclusions that are drawn. When consumers are faced with conflicting information in an area which has previously been characterized by little information, the most natural response is to ignore the new information. This inhibits rather than facilitates promoting the use of performance data for decision making. Thus, finding common methods for drawing summary conclusions from a broad array of data is essential for increasing the likelihood that such information will be used in decision making.

UTILITY OF REPORTS FOR CONSUMERS

How useful was the current report card for consumers? That question is currently being evaluated at the three automobile manufacturers and remains to be answered. But it is worth noting that the process of changing the way people evaluate health plans and make decisions may require several years and multifaceted educational efforts. Evaluation of the current report card is critical, however, for identifying potential improvements in subsequent years. On the other hand, consistency over time may be important for ensuring that the reports get used.

A recent evaluation conducted for NCQA on the utility of report cards for consumers (NCQA and IRE, 1998) reached a number of conclusions that are worth bearing in mind when evaluating the current CARS efforts.

- People who are most likely to remember seeing report cards are those who:
  - tend to use written comparative information in making decisions;
  - have an interest in health and health care;
  - are considering changing health plans.
- The people who find report cards useful are:
  - women (who often make health care decisions in households);
  - those looking to confirm that their current choice is good;
  - those considering changing plans;
  - those with limited experience with the health sector;
  - those with chronic diseases.
- Most report cards have not been shown to have a large impact on consumer decision making, but:
  - a small number of opinion leaders using information may be all that is required to shape market response;
  - choosing among health plans is not a salient activity for everyone, so those most motivated or in need of information may represent the target market (rather than everyone in an eligible population).
- Report cards most likely to be useful when:
  - real differences among health plans exist;
  - choices have consequences;
  - the content is perceived to be relevant;
  - the unit of analysis is meaningful;
  - dissemination strategies match needs.

These points are certainly worth keeping in mind as the collaborative efforts among the automobile manufacturers move forward. And all of these conclusions suggest that a long term, rather than short term view is essential.

HOW USEFUL IS THIS REPORT FOR OTHER PURCHASERS?

Are there attributes of the automobile manufacturers, their employees, or their geographic location that make the current effort unique and not likely to be duplicated with other employers? A fair answer is probably yes and no.

For a variety of reasons, the automobile manufacturers and the UAW have taken a strong leadership position in developing information about quality. Some of this likely stems from the tradition of quality improvement in the manufacturing process. Many of the techniques that are now being used in health care systems to improve care delivery had their beginnings in manufacturing processes. The role of the UAW cannot be

understated. The union represents a large labor force and to conduct collective bargaining for its members in an effective way, there must be some consistency to the methods by which different benefit choices are evaluated. Thus, the pressure to find common ground is reinforced by the UAW which requires such information to facilitate its job. The concentration of the industry in a few states in the Midwest also places the automobile manufacturers in a unique role in their communities—as some of the largest employers they have a significant impact on quality of life in the region. Service to the community is an important part of the tradition of corporate responsibility. Finally, the length of tenure of many employees with these companies is quite long. Thus, the automobile companies are likely to reap the gains of improved quality over the long run—because the workforce will be healthier and more productive. Quality in this case makes good business sense.

Does this mean that other employers will not find this approach useful? Absolutely not. Everyone stands to gain from consistent demands for information about quality and expectations that over time quality will improve. This may be most likely to occur if common methods are used to make judgments—so that there is consistency in determining who is performing well and who requires help improving. It is also important to note that health plans might well serve some populations better than others and that this may make some plans have higher value for some employers than others. Information that both allows consumers to find a plan that meets their needs and that facilitates purchasers setting premiums that offer a fair price is essential to making this scenario work.

WHAT ARE THE LIMITATIONS OF THE CURRENT WORK?

One of the critical limitations of this work is that the evaluations are applied to just one sector of the health delivery system—managed care plans. In fact, two-thirds of the automobile manufacturers beneficiaries are currently covered by traditional indemnity insurance (TRAD) or preferred provider organizations (PPO). Failing to provide information about the performance of the TRAD sector creates several problems. First, this may send a signal that performance in the traditional sector is optimal and requires no monitoring to identify opportunities for improvement. This conclusion is inconsistent with a body of research literature that finds few differences between managed care and fee-for-service delivery. Second, monitoring only the managed care sector may impose unfair costs of doing business on that sector since plans must spend money to collect data on quality but the indemnity sector is not required to do so. Third, it may be confusing to

consumers to have information on a subset of their health insurance choices. This may cause people to disregard the information on quality or misinterpret it.

The information presented in this report is based on the data that are currently publicly available. Certainly everyone agrees that there are a number of areas in which more information is required. For example, there are considerably fewer measures available on the quality of care for acute conditions or management of chronic illnesses than there are for preventive services. Much more information is available on the quality of care for women than men, and adults than children. NCQA and others are working to fill these gaps, but this will take time. In the meantime, we may have to make do with a less than perfect set of information for some important areas.

More sophisticated research methods could have been used for some of the key decisions. For example, we could have used regression imputation methods to more closely estimate actual performance of plans with missing data on certain measures. Other approaches to weighting individual measures or to combining measures into overall scores are possible. While it is certainly worth exploring whether these methods would produce different results, there are challenges in determining what represents the best method. We have no real "gold standard" against which summary scores can be compared to determine their predictive value. The principal method available today is face validity (does the method make sense, do the "right" plans seem to do well or poorly on the final scores). But this is somewhat circular—it presupposes that we already know who the winners and losers are and that if the facts don't conform to those presuppositions then the facts or methods are wrong. So, continued evaluation of the methods is important but difficult.

WHAT ARE THE NEXT STEPS?

At the beginning of the project, it was agreed that this process was an iterative one that would take place over several years. Establishing a quality improvement cycle that allows each subsequent iteration to address problems encountered in the last iteration should facilitate development of a system that optimizes the use of available information and that perhaps enhances the likelihood that more information will be made available.

We have proposed in the coming year that some key issues be addressed:

- Incorporating reporting on TRAD
- Investigating the contribution of stronger scientific methods
- Increasing the number of scoring categories

- Obtaining feedback on the utility of current reports (and suggestions for how reports could be made more useful)

CARS '98 represented a substantial contribution to making data useful for decision making. The collaborative nature of the project—bringing together competitive employers, union representatives, public purchasers, business groups and researchers to produce consensus and implement the decisions in a four month time frame is quite remarkable. It is certainly well worth continuing to build on this foundation.

APPENDIX A
SAMPLE VERIFICATION REPORT

Health Plan Name	**Health Plan A**
NCQA Organization ID	**XXX**
Childhood Immunization Rate (Combo 1)	71.29
Adolescent Immunization Rate (Combo 1)	8.76
Prenatal Care in the 1st Trimester	71.76
Breast Cancer Screening	77.18
Cervical Cancer Screening	82.35
Check-up After Delivery	70.22
Well-child visits: 0-15 months (0 visits)	4.64
Well-child visits: 3-5 years	55.13
Adolescent Well-Care Visits	33.25
Follow-Up for MH Hospitalization	71.57
Eye Exams for Diabetics	35.87
Beta-Blocker Treatment after Heart Attack	

Health Plan Name		**Health Plan A**
NCQA Organization ID		**XXX**
7a. Ease of making appointments for medical care by phone	no experience	16
	poor	33
	fair	75
	good	199
	very good	219
	excellent	184
	missing	7
7b. Length of time you had to wait between making an appointment for routine care and the day of your visit	no experience	20
	poor	67
	fair	109
	good	220
	very good	1
		98
	excellent	107
	missing	12
7c. Thoroughness of treatment you received	no experience	17
	poor	20
	fair	68
	good	195
	very good	261
	excellent	162
	missing	10
7d. Attention given to what you had to say	no experience	18
	poor	20
	fair	82

	good	190
	very good	239
	excellent	177
	missing	7
7e. Number of doctors you had to choose from	no experience	76
	poor	45
	fair	83
	good	213
	very good	189
	excellent	102
	missing	25
7f. Ease of choosing a personal physician	no experience	52
	poor	37
	fair	85
	good	216
	very good	188
	excellent	132
	missing	23
7g. Amount of time you had with doctors and staff during a visit	no experience	22
	poor	42
	fair	90
	good	226
	very good	230
	excellent	114
	missing	9
7h. How much you were helped by the care you received	no experience	22
	poor	20
	fair	77
	good	228
	very good	250
	excellent	128
	missing	8
7i. Overall quality of care and services	no experience	18
	poor	13
	fair	69
	good	220
	very good	272
	excellent	136
	missing	5
8a. Types of services the plan covered	no experience	38
	poor	24
	fair	79
	good	234
	very good	228
	excellent	116
	missing	14
8b. Availability of information from your plan about eligibility; covered services or administrative issues	no experience	55
	poor	35
	fair	81
	good	257
	very good	186
	excellent	98

	missing	21
8c. Availability of information from your doctors or plan about costs of care	no experience	101
	poor	43
	fair	91
	good	220
	very good	171
	excellent	81
	missing	26
8d. Length of time you had to spend filling out claim forms or other paperwork	no experience	136
	poor	14
	fair	37
	good	138
	very good	188
	excellent	205
	missing	15
9a. Delays in your medical care while you waited for approval by your health plan	big problem	36
	small problem	79
	not a problem	585
	missing	33
9b. Difficulty in receiving care you and your doctor believed was necessary	big problem	42
	small problem	72
	not a problem	585
	missing	34
9c. Not being able to get a referral to a specialist that you wanted to see	big problem	71
	small problem	81
	not a problem	538
	missing	43
10. Have you called or written your health plan with a complaint or problem in the last 12 months?	yes	126
	no	588
	missing	19
10a. How long did it take for the health plan to resolve your complaint?	same day	37
	1 week	14
	2 weeks	6
	3 weeks	3
	4 weeks	24
	not yet	40
	missing	2
11. All things considered, how satisfied are you with your current health plan?	completely satisfied	141
	very satisfied	317
	somewhat satisfied	139
	neither satisfied nor dissatisfied	50
	somewhat dissatisfied	42
	very dissatisfied	19
	completely dissatisfied	8
	missing	17

13. Would you recommend your current health plan to your family or friends if they needed care?	definitely yes	218
	probably yes	401
	probably not	88
	definitely not	14
	missing	12
14. Do you intend to switch to a different health plan when you next have an opportunity?	definitely yes	22
	probably yes	62
	probably not	406
	definitely not	218
	missing	25
15a. Routine care	same day	19
	1-3 days	112
	4-7 days	121
	8-14 days	110
	15-30 days	118
	31-60 days	94
	61 days or longer	72
	no experience	63
	missing	24
15b. Minor illness or injury	same day	166
	1-3 days	320
	4-7 days	61
	8-14 days	20
	15-30 days	7
	31-60 days	0
	61 days or longer	0
	no experience	127
	missing	32
15c. Chronic or ongoing condition	same day	101
	1-3 days	165
	4-7 days	84
	8-14 days	32
	15-30 days	21
	31-60 days	14
	61 days or longer	5
	no experience	256
	missing	55
15d. Urgent care	same day	341
	1-3 days	37
	4-7 days	6
	8-14 days	0
	15-30 days	2
	31-60 days	0
	61 days or longer	0
	no experience	289
	missing	58
16. Over the last 12 months, when calling for medical information or advice, how long did it usually take for your provider's offices to return your call?	no experience	236
	less than 1 hour	72
	1 hour but less than 4 hours	205

	4 hours but less than 7 hours	113
	7 hours but less than 24 hours	41
	24 hours or more	42
	missing	24
17. Over the last 12 months, once you got to your provider's office, how long did you usually have to wait to see your provider when you had an appointment for care?	no experience	36
	less than 10 minutes	72
	10-15 minutes	275
	16-30 minutes	226
	more than 30 minutes but less than 45 minutes	52
	45 minutes to 1 hour	39
	1 to 2 hours	11
	2 hours or more	2
	missing	20
18. Over the last 12 months, when you went for medical care, how often did you see the same provider?	no experience	33
	always	405
	most of the time	224
	sometimes	43
	rarely or never	11
	missing	17
32a. On how many of these visits were you advised to quit smoking?	no experience	23
	1 visit	40
	2 to 4 visits	33
	5 to 9 visits	9
	10 or more visits	6
	missing	4