

2. EVALUATION METHODS

The evaluation of CRI was designed to use information from the demonstration sites and from a set of matched control sites during two time periods: before CRI began and after CRI was fully implemented. Because the demonstration was conducted without random assignment of families to different health care plans, it could not be considered an experiment. It was, however, a rare opportunity to look at the effects of health care options in a “natural experiment.”

The evaluation encompassed ten military hospital catchment areas and one military clinic area¹ within the demonstration states as well as ten military hospital catchment areas and one clinic area that served as comparison or control sites (Table 1). The control sites were located outside of California and Hawaii, and did not participate in the demonstration. The control sites were matched individually to the demonstration sites on the basis of service affiliation, hospital size, CHAMPUS population size by beneficiary type, utilization of MTF and CHAMPUS care, and the ratio of MTF to CHAMPUS utilization. The methods used to select the demonstration and control sites are described in Sloss et al. (forthcoming).

Table 1
Demonstration and Control Sites for the RAND
CRI Evaluation Project

Service	Demonstration Site	Control Site
Army	Ford Ord, CA	Fort Hood, TX
	Tripler, HI	Madigan, WA
Navy	Long Beach, CA	Orlando, FL
	Camp Pendleton, CA	Charleston, SC
	San Diego, CA	Portsmouth, VA
	Port Hueneme, CA (clinic)	Quantico, VA (clinic)
Air Force	Beale, CA	Dover, DE
	March, CA	Carswell, TX
	Mather, CA	Homestead, FL
	Travis, CA	Keesler, MS
	Vandenberg, CA	Shaw, SC

¹The primary difference between a military hospital and a military clinic is that a clinic provides no inpatient care. Some military clinics provide a volume of outpatient care comparable to a military hospital.

The design of the evaluation allowed us to measure the demonstration's effects by comparing the demonstration sites with the control sites before and during the demonstration period. By observing the demonstration sites at two points in time (before and during CRI), we estimated the change that occurred over time. To determine whether this was different from what would be expected in the absence of CRI, the demonstration sites were compared with the control sites. The control sites provided a measure of how the demonstration sites might have been expected to fare if CRI had not been implemented. The experience of the control sites also reflected the effects of any systemwide policy changes. If changes not related to CRI occurred in one set of sites, but not the other, the effects of those changes could not be separated from the CRI effect. However, our on-site interviews have not uncovered any notable differences between the demonstration and control sites other than the CRI program (Anderson et al., forthcoming).

The effects of the demonstration documented in this report correspond to two of the demonstration's goals: to improve access to care and satisfaction with care among CHAMPUS beneficiaries. The rest of this section describes the data used in these analyses, as well as the methods employed in evaluating and analyzing the measures that are used to determine whether the demonstration met its goals.

MEASURING ACCESS AND SATISFACTION

This analysis attempted to determine if the level of beneficiary access or satisfaction increased or decreased more among those living in the demonstration sites than in the control sites. The analysis compared access and satisfaction two years after CRI began between CHAMPUS beneficiaries in the demonstration sites (members of CHAMPUS Prime and those not enrolled, separately) and CHAMPUS beneficiaries in the control sites. We assumed in this comparison that the control sites represent what the demonstration sites would have experienced if CRI had not been introduced. The difference between the experience in the demonstration sites and in the control sites was therefore attributed to the effect of CRI.

Through our statistical methods, we controlled for two types of differences that might otherwise mask the true effect of CRI. First, because individual characteristics can affect the access and satisfaction reported by a person, we control for personal health and sociodemographic characteristics. Second, because other systemwide changes can occur within the Military Health Services System (MHSS) at the same time as CRI, we cannot attribute a measured effect on access or

satisfaction to CRI without accounting for these changes. We control for MHSS systemwide changes by analyzing satisfaction in the control sites. In particular, using regression analysis, we compare satisfaction before and during CRI in the demonstration sites with that observed in the control sites.

A third type of change may occur. Changes that are unrelated to the demonstration—new programs, new systems, or new rules—may be introduced at individual sites. Regression analysis cannot be used to account for these site-specific changes. If changes occur in the demonstration sites during the same time period that affect access or satisfaction, they will be inextricably tied to the effects of the demonstration. Changes at other individual sites may also occur during this time period. Any site-specific changes will be documented in Anderson et al. (forthcoming). They will not be incorporated into our analysis in any quantitative or direct way, but they may offer possible explanations for results observed in the evaluation.²

EXPECTED EFFECTS OF CRI PROGRAMS ON ACCESS AND SATISFACTION

Many of the changes implemented as part of CRI would be expected to have an effect on access and satisfaction. Measuring the separate effect of each individual change is impossible because all of the changes occurred in all sites at essentially the same time. Therefore, we measured the effect of CRI as a whole rather than the effect of any one change.

Although increasing access to care is a goal of CRI, beneficiaries' perception of CRI's effect on access is not necessarily straightforward. Although many aspects of CRI are designed to improve access, other changes might be perceived as less than desirable. The Health Care Finder service is designed to improve access by helping all beneficiaries (CHAMPUS Prime members and others) schedule referrals and find appropriate providers in the military or civilian sectors. The effect of the Health Care Finder service should be more pronounced among those enrolled in CHAMPUS Prime. Almost 60 percent of the calls handled by the Health Care Finder in CRI catchment areas offering CHAMPUS Prime were attributable to the 13 percent of beneficiaries enrolled.³ Although many of the changes were targeted at

²These explanations will be explored in Volume 1 of this report series (Hosek et al., forthcoming).

³These figures are taken from the Annual Health Care Finder Referral Report for Option Period 4, submitted by Foundation Health Corporation.

improving access to civilian care, CRI also may have improved access to appointments in the MTF. In addition, better coverage of preventive care services under CHAMPUS Prime would be interpreted as an improvement in benefits to most, and probably would increase use of (access to) these services. On the other hand, however, CHAMPUS Prime members are required to use physicians and hospitals in the contractor's network, meaning they do not have complete freedom in choosing their provider. In addition, network physicians and hospitals might be less conveniently located than providers outside the network.

In general, CRI would be expected to increase satisfaction levels among beneficiaries. The enrollment program, CHAMPUS Prime, offers civilian care to its members at a lower out-of-pocket cost than standard CHAMPUS, whereas CHAMPUS Extra is intermediate in terms of beneficiary cost. As a result, one would expect satisfaction with the cost of civilian care to increase substantially among CHAMPUS Prime members and, to a lesser extent, among CHAMPUS Extra users. However, CRI's effect on satisfaction with other aspects of civilian and military care might be less predictable. As mentioned above, many of the changes under CRI are aimed at improving access, but beneficiaries might not be satisfied with all of the changes, and therefore might not be satisfied with access. Satisfaction with doctor care is determined largely by the individuals who provide the care, rather than by the system in which the care is provided. Therefore, predicting whether CRI would affect it in a positive or negative way is difficult.

The general framework used to evaluate access and satisfaction leads to the testing of two hypotheses for each outcome measure:

- Is there a difference under CRI between those enrolled in CHAMPUS Prime and those in the control sites?
- Is there a difference under CRI between those in the demonstration sites not enrolled in CHAMPUS Prime and those in the control sites?

We analyzed the survey measures using multivariate regression methods. We assessed differences between each of the two demonstration groups and control sites at the time of the follow-up survey, controlling for levels measured in the baseline survey.

SURVEY ADMINISTRATION

All of the outcomes related to access and satisfaction are based on data collected in a survey of CHAMPUS beneficiaries. CHAMPUS beneficiaries include dependents (spouse and children) of active-duty military members, retired military members and their dependents, and survivors of military personnel.⁴ We administered the survey twice, once before CRI was implemented (baseline) and two years after CRI began (follow-up). The baseline survey was administered to CHAMPUS beneficiaries living in the demonstration and control sites during the summer and fall of 1988 before the CRI programs were fully implemented. In the fall and winter of 1990–1991, over two years after services under CRI were first offered in California and Hawaii, the follow-up survey was administered to a different set of beneficiaries living in the same catchment areas. There was no overlap between the two survey samples; that is, we excluded from the follow-up sample all people who had been selected as part of the baseline sample. We thought that people who had received the baseline questionnaire might answer questions differently or might be more (or less) likely to complete a questionnaire than those who had not received a baseline questionnaire, resulting in possibly biased information.

The survey samples were drawn so that the surveys completed by active-duty families would represent a random sample of all active-duty families in the study sites, and similarly the surveys completed by retiree families would be a random sample of all retiree families. The distribution of the sample by catchment area was proportional to the total beneficiary population living in the areas at the time of the surveys. We determined sample sizes on the basis of our ability to detect a difference of 5 percent of the standard deviation in outpatient visits—about 0.3 visits—for active-duty and retired families separately, with the power set at 0.80⁵ and the significance level set at 0.025.⁶ We can detect much smaller differences for most of the outcomes evaluated in this report. We oversampled CHAMPUS Prime enrollees in the follow-up survey to allow separate analyses of their

⁴Eligibility for CHAMPUS benefits ends at age 65, so all survey respondents were under age 65.

⁵This means there is an 80 percent probability of detecting a difference when in fact there is a difference.

⁶This means there is a 2.5 percent probability of finding a difference when in fact there is no difference.

data.⁷ For a more detailed description of the sample design, see Sloss et al. (forthcoming).

The baseline questionnaire consisted of mostly multiple-choice questions on health status, use of health care in military and civilian systems, use of preventive care, inability to get care when needed, satisfaction with care, experience at a recent medical visit, other insurance coverage, demographic and socioeconomic characteristics, and military service. Many questions were selected from previous studies of similar topics.⁸ We added questions to the follow-up survey on the two CRI programs, CHAMPUS Prime and CHAMPUS Extra.

The respondents received the questionnaires through the mail, along with a letter describing the study and encouraging participation. Each family in the sample was sent one questionnaire to be completed by an adult. If the family had one or more children 21 years of age or under, a child questionnaire was also sent. In each active-duty household, the spouse of the active-duty military member was always the respondent.⁹ In each retiree household with two CHAMPUS-eligible adults (retiree and spouse), one was randomly selected as the respondent.¹⁰ In households with more than one child, one child was selected randomly as the respondent.

In the baseline survey, we mailed over 12,000 surveys to active-duty military families, and almost 11,000 to families of retired military members (Table 2). At follow-up, almost 14,000 surveys were mailed to active-duty families and over 11,000 to retiree families. The results for adults presented in this report are based on 11,183 baseline questionnaires (4724 active-duty spouses and 6459 retirees or retiree spouses) and 11,138 follow-up questionnaires (4,802 active-duty spouses and 6336 retirees or spouses). The results for children are based on 4288 baseline questionnaires (2947 from active-duty

⁷Follow-up surveys were mailed to 1912 active-duty families and 1203 retiree families who were enrolled in CHAMPUS Prime. These sample sizes represent an oversampling rate of 48 percent and 32 percent, respectively, over the number of CHAMPUS Prime families that were drawn in the initial sample. Our goal was to receive at least 1500 surveys completed by families enrolled in CHAMPUS Prime, divided equally between active-duty and retiree.

⁸Many of the questions related to health status and health perceptions are based on a short-form health survey used in the Medical Outcomes Study (Stewart et al., 1988).

⁹Active-duty military members were not included in the survey sample because they are not eligible for CHAMPUS benefits.

¹⁰Households with no CHAMPUS-eligible adults (retiree or spouse) were not included in the survey sample.

Table 2
Breakdown of Survey Sample by Response Category

	Baseline		Follow-up	
	Number	Percent	Number	Percent
Total				
Surveys mailed	23,036	—	24,975	—
Effective sample size ^a	17,867	100.0	20,362	100.0
Completed questionnaires	11,183	62.6	11,138	54.7
No response	6,684	37.4	9,224	45.3
Active duty families				
Surveys mailed	12,355	—	13,833	—
Effective sample size ^a	8,198	100.0	10,185	100.0
Completed questionnaires	4,724	57.6	4,802	47.2
No response	3,474	42.4	5,383	52.8
Retiree families				
Surveys mailed	10,681	—	11,142	—
Effective sample size ^a	9,669	100.0	10,177	100.0
Completed questionnaires	6,459	66.8	6,336	62.3
No response	3,210	33.2	3,841	37.7

^aExcludes three types of surveys: those not delivered because of wrong or incomplete addresses, those determined to be ineligible, and those completed by people who had moved out of the study area.

families and 1341 from retiree families) and 4786 follow-up questionnaires (3063 from active-duty families and 1723 from retiree families).

Response rates were fairly low. Only 67 percent and 62 percent of retiree families and 58 percent and 47 percent of active-duty families completed the questionnaires at baseline and follow-up, respectively (Table 2).¹¹ Response rates were considerably higher among families of officers than enlisted personnel, and consistently highest among Air Force families. Active-duty families with children were also more likely to respond, whereas retirees with children were less likely. For a more detailed discussion of response rates in the CRI beneficiary surveys, see Sloss et al. (forthcoming).

ACCESS MEASURES

Many aspects of medical care can be considered measures of access, including utilization, system characteristics preventing utilization, and patient satisfaction. We limited our access measures to those we

¹¹The larger decrease in response rate between baseline and follow-up for active-duty families is probably tied to the deployments to the Persian Gulf in the last three months of 1990, decreasing response to repeat mailings.

felt were the most important and at the same time the most easily obtained through a self-administered questionnaire. We used a subset of the measures of “potential” and “realized” access described by Aday et al. (1984). For respondents to answer the questions related to access required little or no subjective judgment, because the information requested can be recalled objectively and easily.

“Potential” access measures are “indicators that describe the process of obtaining care” (Aday et al., 1984). Measures of “potential” access relate to the process of obtaining care and aspects of the system that might interfere with that process. We examined three measures of potential access: the existence of a regular source of care, waiting time in a doctor’s office, and number of telephone calls needed to schedule an appointment. Previous studies indicated that having a regular or usual source of care has an important effect on how much care is used and how satisfied one is with the care received (Aday et al., 1984; Aday and Andersen, 1975). The length of time people wait to see a doctor is another measure of potential access that reflects on the operating efficiency of the system (Aday et al., 1984), as well as being a frequent target of criticism by patients in the MTFs. Including number of telephone calls needed to schedule an appointment as a measure is also aimed at identifying barriers to access, similar to office waiting time. This aspect of military care is also often criticized. Because beneficiaries and health care administrators alike consider the inconveniences encountered in getting care as an indication that problems exist in accessing the system, it is important to use these measures to evaluate how efficiently the beneficiaries can access the system.

“Realized” access measures are defined as “services actually received in terms of units of care” (Aday et al., 1984). Under this category, we have included the percentage of beneficiaries with an outpatient visit in the past six months as a measure of their realized ability to get into the system. Another measure of “realized” access (or perhaps “unrealized” access) is whether a respondent needed to go for medical care, but did not. For these two measures, we repeated the analysis for adults twice: first including everyone, and then including only those who reported having a medical condition for which effective treatment is available. These “highly treatable” conditions included: chronic bronchitis, asthma, emphysema, phlegm production, severe lung problem, chest pain, angina, heart attack, heart failure, enlarged heart, hypertension, diabetes, and cancer (Lohr et al., 1986).

We also queried respondents about their use of preventive care as another indicator of demonstrated or “realized” access. Because the en-

hanced benefit package offered to CHAMPUS Prime members includes improved coverage of preventive care, we wanted to evaluate whether the members are taking advantage of the opportunity to undergo screening tests and examinations. We asked each female respondent when was her most recent experience with a gynecologic examination with Pap smear, a breast examination by a physician, a mammogram, and prenatal care (for those pregnant within the past 12 months). In addition, we asked all respondents when their most recent general physical examination, blood pressure check, and rectal examination was.

The survey questions on which these measures of access are based are reproduced in Appendix A.

SATISFACTION MEASURES

The first of two types of satisfaction, the *general satisfaction* scales, focuses on attitudes toward medical care in general, rather than on any particular episode or encounter. The second type consists of seven scales measuring *encounter-specific satisfaction* based on ratings of the most recent medical visit.

General Satisfaction

Four multi-item satisfaction scales allow us to score the respondent's overall attitude toward medical care and three specific aspects of care: cost, the interaction between the doctor and patient, and the ability to get care. The questions on which the scales are based require the respondent to provide a subjective opinion. Most of the questions consist of brief statements describing attitudes toward medical care, followed by a five-point scale on which the respondent indicates level of agreement. The general satisfaction items used in this study have been used in the same or a similar form in several previous studies (Ware et al., 1976a,b; Ware et al., 1983; Davies et al., 1986). Studies using these satisfaction measures have shown that there can be behavioral consequences of decreased satisfaction. One study showed that a decrease in satisfaction predicted subsequent changes in medical care providers (Marquis et al., 1983), whereas another showed that low satisfaction predicted disenrollments from prepaid health plans (Ware and Davies, 1983).

The general satisfaction scales are based on 18 items from a 50-item satisfaction battery used in the patient questionnaire from the Medical Outcomes Study (Tarlov et al., 1989). To decrease the respondent burden, the 50-item battery was shortened to 18 items on

the basis of a psychometric evaluation designed to retain a subset representing all the dimensions of satisfaction present in the original battery. The scales based on the 18-item battery were found to correlate highly with the scales based on the original 50-item battery (see Appendix B).

Table 3 shows the four general satisfaction scales and the items within each scale. Respondents chose a response to each item, using a five-point scale (strongly agree to strongly disagree). The scale is formed by summing the response values of all items in that scale.¹² Nine of the 18 items were reversed before the items were summed (marked with * in Table 3). For example, a higher score (stronger disagreement) on item 17 indicates higher satisfaction with medical care in general, whereas a higher score on item 3 indicates lower satisfaction. Before these items were combined to form the general satisfaction scale, the responses to item 3 were reversed in value, so a higher score on the scale indicates higher satisfaction. We then transformed the summed value to a 0-to-100 scale¹³ using the following formula:

$$\text{Scale value} = \frac{S - \text{MIN}}{\text{MAX} - \text{MIN}} \times 100$$

S = sum of items

MIN = minimum possible scale value

MAX = maximum possible scale value

We calculated scale values for every individual who had responded to one or more of the items comprising the scale. A missing value for a scale was assigned only when the individual was missing all items in the scale.

We performed extensive psychometric evaluation of the general satisfaction scales, using data from the baseline and follow-up CRI surveys. We evaluated the internal consistency of the scales using Cronbach's (1951) alpha coefficient. We also reviewed the item-scale correlation matrix for item convergence and item discrimination for each scale.

¹²This summated-rating scale is often referred to as a Likert-type scale.

¹³This standardization facilitates comparison of values among scales based on different numbers of items.

Table 3
Questionnaire Items in General
Satisfaction Scales

Scale Name and Item Content	Order of Items in Questionnaire
Overall satisfaction	
Care just about perfect*	3
Dissatisfied with some things	17
Access	
Easy access to specialists*	8
Wait too long for emergency treatment	9
Hard to get appointment right away	16
Get medical care whenever needed*	18
Financial aspects	
Care without financial setback*	5
Pay more than I can afford	7
Doctor care	
Explain the reason for tests*	1
Office has everything needed*	2
Wonder if diagnosis is correct	4
Careful to check everything*	6
Too business-like, impersonal	10
Very friendly and courteous*	11
Hurry too much when treating me	12
Ignore what I tell them	13
Doubt about ability of doctors	14
Doctors spend plenty of time*	15

*Item scores were reversed before scale was constructed.

The magnitude of Cronbach's (1951) alpha coefficients suggested the four general satisfaction scales were internally consistent¹⁴ and supported the construction of the hypothesized scales. The relatively high correlations between the items and their hypothesized scales also indicate item convergence.¹⁵ In addition, the fact that the correlation between each item and its scale was generally higher than the correlations between the same item and other scales demon-

¹⁴Based on Helmstadter's (1964) recommendations that scales with internal-consistency (alpha) estimates of 0.50 or above are appropriate for group comparisons.

¹⁵A corrected correlation of 0.40 or higher has been recommended as an indicator of item convergence for scales being newly developed, whereas higher correlations should be required for previously used scales (Hays and Hayashi, 1990).

strates item discrimination.¹⁶ The alpha coefficients and item-scale correlations are given in Appendix C.

Two aspects of the analyses of the general satisfaction scales differ from the methods used to analyze the other access and satisfaction measures. First, the active-duty families were analyzed separately from the retired families. Second, each respondent was asked to rate the system of care he was “most familiar with,” either military or civilian. As a consequence, we also analyzed military and civilian ratings separately.

The survey questions on which the general satisfaction scales are based are reproduced in Appendix A (see question 48).

Encounter-Specific Satisfaction

Seven scales address the respondent’s attitude toward his or her most recent medical encounter within the past six months. The topics covered by the items include satisfaction with the visit overall and with specific aspects of the visit: travel time, office waiting time, length of visit, information about condition and treatment, cost, and quality of care. These questions have been adapted from similar questions asked as part of the 1982 National Survey of Access to Medical Care (Aday et al., 1984).¹⁷

Each of these encounter-specific satisfaction scales is based on a single item rated on a 1-to-5 scale, ranging from “very satisfied” to “very dissatisfied.” These scales were also transformed to a 0–100 scale, using the method described for general satisfaction above. In the regression analyses of these scales, we have excluded walk-in visits and have analyzed military and civilian appointments separately.

The survey questions on which the encounter-specific satisfaction scales are based are reproduced in Appendix A (see question 41).

METHODS FOR ANALYZING ACCESS AND SATISFACTION

Our first set of analyses compared simple mean values from the follow-up survey for CHAMPUS Prime members and others living in the

¹⁶The items in the overall satisfaction scale also correlated highly with two other scales (doctor care and access). However, because the overall scale is a summary measure, the item discrimination criteria (and collapsing of scales) do not apply.

¹⁷This telephone survey was conducted by Lou Harris and Associates, with support from the Robert Wood Johnson Foundation, on a sample of 4800 families in the United States.

demonstration sites, and the control sites. After this simple analysis, we estimated the differences between the demonstration sites and the control sites using multiple linear regression (ordinary least-squares or OLS method) or multiple logistic regression. In these regression analyses, the model identifies baseline respondents who later joined CHAMPUS Prime (based on information from the CRI contractor). The regression model statistically controls for differences between them and other baseline respondents who did not join. Because of this analytic approach, the differences between the follow-up values (from the regressions) should represent the effect of CRI, and should not be confounded by the self-selection of a certain type of beneficiary into CHAMPUS Prime. For a more detailed discussion of the statistical methods, see Appendix E.