May 2006

Evaluation of the Cancer Prevention and Treatment Demonstration

Final Design Report

Prepared for

Arthur Meltzer, Ph.D.
Centers for Medicare & Medicaid Services
DHHS/CMS/OA/ORDI/REG/DBR
Mail Stop C3-19-25
7500 Security Boulevard
Baltimore, MD 21244-1850

Prepared by

Janet B. Mitchell, Ph.D.
Susan G. Haber, Sc.D.
Deborah Holden, Ph.D.
Sonja Hoover, M.P.P.
Kevin Smith, M.A.
Sujha Subramanian, Ph.D.
RTI International
411 Waverley Oaks Road
Suite 330
Waltham, MA 02452-8414

RTI Project Number 0207964.027

EVALUATION OF THE CANCER PREVENTION AND TREATMENT DEMONSTRATION

Final Design Report

Authors: Janet B. Mitchell, Ph.D.
           Susan G. Haber, Sc.D.
           Deborah Holden, Ph.D.
           Sonja Hoover, M.P.P.
           Kevin Smith, M.A.
           Sujha Subramanian, Ph.D.

Project Director: Janet B. Mitchell, Ph.D.

Federal Project Officer: Arthur Meltzer, Ph.D.

RTI International*

CMS Contract No. 500-00-0024 TO#27

May 2006

---

*RTI International is a trade name of Research Triangle Institute.

CONTENTS

List of Figures

List of Tables

SECTION 1
INTRODUCTION

The Centers for Medicare & Medicaid Services (CMS) contracted with RTI International in 2005 to evaluate the Cancer Prevention and Treatment Demonstration (CPTD). The purpose of the CPTD is to reduce disparities in screening, early detection, and treatment of cancer among ethnic and racial minorities who are Medicare beneficiaries. The following Section provides a brief overview of the literature describing racial and ethnic disparities in screening and treatment for various types of cancer and an overview of the demonstrations.

1.1 Statement of the Problem

Cancer mortality has been declining for most populations in the United States, but this benefit is less obvious for some racial and ethnic groups, as well as for others defined by geographic region or income. As an example, African Americans are more likely to develop and die from cancer than persons of any other racial and ethnic group (ACS, 2004). Through a closer examination of the literature and of statistical trends, it is clear that the benefits of current knowledge about state-of-the-art cancer care are not shared equally by all members of our society (Aziz and Rowland, 2002) and reflect a disconnect between research discovery, system development, and service delivery (NIH/NCI, 2001). Some reasons for these disparities are known, while others are still under study,. The known causes include limited access to health care (e.g., increasing numbers of uninsured Americans, decreasing resources for Medicaid coverage, geographic location of health facilities), differences in medical practices provided to patients, lack of knowledge among people about the types of care that are of greatest benefit, and others.

1.1.1 Screening

Identifying and eliminating the barriers to cancer control and treatment is a necessary step to achieving the Healthy People 2010 and the National Cancer Institute (NCI) Director's 2015 goals of reducing these disparities (USDHHS, 2003, 2004; Harper and Lynch, 2005). Enhanced quality of life and longevity after a cancer diagnosis are influenced by the extent to which the cancer has metastasized prior to diagnosis. This factor is greatly influenced by the adequacy of care received and whether a screening procedure is provided early enough so that cancer is found when it is still localized. All the cancers, with the exception of those of the lung, (i.e., prostate, breast, cervical, colorectal) for CMS' CPTD have existing screening tests enabling early detection.

Even though these tests are available, minorities and people from low socioeconomic backgrounds are less likely than whites to receive cancer screening services and more likely to have late-stage cancer when the disease is diagnosed, except for the case of African American women who have higher screening rates for cervical cancer than whites or Asians (AHRQ, 2004). Despite the higher screening rates for cervical cancer among African American women, their mortality rate from this disease is the highest, at 5.9/100,000 women (age-adjusted), as compared with 3.7/100,000 among Hispanic women and 2.7/100,000 deaths attributable to cervical cancer among White women (ACS, 2004). In addition, among the elderly, African Americans were less likely than Whites to undergo the prostate-specific antigen (PSA) screening (Gilligan, et al., 2004). Elderly African Americans were also found to have fewer screening tests

for colorectal cancer compared to elderly Whites (Cooper and Koroukian, 2004). Similarly, other racial and ethnic minority groups, not just African Americans display levels of cancer screening below those of the White population (NIH/NCI, 2001).

1.1.2 Treatment

Differences in primary treatment of breast, cervical, colorectal, lung, and prostate cancers, as well as appropriate adjuvant therapy, have all been documented to exist between whites and various minority populations (AHRQ, 2004). While differences in cancer mortality rates can reflect a variety of factors such as genetic disposition and lifestyle, screening and early treatment can lead to significant reductions in mortality, particularly for breast and cervical cancer. African Americans have the highest death rate from all cancer sites combined (ACS, 2004), and the 5-year survival rates across all racial and ethnic groups is more than 10% higher for persons who live in higher SES areas of the country. Even when local poverty levels are controlled, African Americans, American Indian and Alaska Native (AI/AN), and Asian/Pacific Islander (A/PI) and A/PI men and African American and AI/AN women have lower 5- year survival rates than Whites (ACS, 2004).

Numerous studies have focused on racial and ethnic disparities in the survival rates of Medicare beneficiaries using linked Medicare claims and SEER (Surveillance, Epidemiology, and End Results) data. In addition, several recent studies have examined racial and ethnic differences in cancer treatment using these linked data. One study found that although African Americans and Whites were equally likely to consult with medical oncologists, African Americans were less likely than Whites to undergo chemotherapy for colon cancer treatment (Baldwin, et al., 2005). Some of the disparity could be explained by patients' severity of illness and length of stay, whose underlying causes could be complications, health status, and level of home care support. In addition, lower educational levels were indicative of lower chemotherapy rates.

In separate studies using Medicare and SEER data, African Americans were also less likely than Whites to undergo surgery for lung cancer (Lathan, et al., 2006; Bach, et al., 1999). Further, Lathan et al. found that African Americans were less likely than Whites to have surgery recommended to them, which was also suggested by Bach, et al., and Blacks were more likely to refuse it. Authors of both studies suggested additional research in this area, particularly qualitative. Similar racial differences have been documented in the area of cardiac surgery (e.g., Cromwell et al., 2006).

1.1.3 Barriers

There are numerous barriers to obtaining any of the cancer screening tests, particularly among un- or under-insured populations and/or those living in rural areas of the country. Common barriers include:

- Limited or no access to a provider on a regular basis (AHRQ, 2004);

- Poor communication with providers, particularly for non-English speaking populations;

- Lack of knowledge of the need for ongoing cancer screening (vs. a one time test) (Ogedegbe, et al., 2005);

- Perceptions of 'feeling good' so thinking nothing could be wrong (Ogedegbe, et al., 2005);

- Lack of access either because of operating time of the clinics (no after hours offered), lack of transportation or limited child care;

- Fear of the cancer test itself and/or of finding cancer; and,

- Lack of a recommendation from a clinician.

The recognized effectiveness of an organized cancer screening program is undermined if follow-up of abnormal screening results is delayed or never completed. Barriers to follow-up include limited access to necessary specialized services and a lack of understanding of the meaning of a test result and the importance of following provider recommendations.

Seeking treatment presents similar barriers. Studies have shown that racial and ethnic groups are less likely than whites to receive post-treatment surveillance and even certain types of treatment options (AHRQ, 2004). Examples include a study where African American patients with metastasized cancer were 27% less likely than whites to receive a major treatment (Bach et al., 1999), and African Americans had a 63% greater probability of being untreated for pain relative to whites (Bernabei, et al., 1998).

1.1.4 Interventions

Research over the years has shown that successful interventions incorporate various levels of influence, such as the interpersonal, community, and organization levels, into initiatives that focus on health determinants (IOM, 2000: Holden, et al., 1998). One of the most important means of overcoming barriers faced by racial/ethnic minorities and of improving cultural competency is enlisting community health workers to assist in educating and promoting screenings and treatment (Wolff, et al. 2003; Brandeis 2003). Several recent studies have indicated the need for more research to determine whether patient navigators, community health workers who steer patients through the health care system, help to reduce barriers among the underserved (Dohan and Schrag 2005, Hede 2006). The CMS CPTD will begin to address many of these issues, particularly among the most vulnerable populations of elderly minority patients.

1.2 Overview of the Demonstration

Section 122 of the Medicare, Medicaid, and SCHIP Benefits Improvement and Protection Act (BIPA) of 2000 stipulated that the Secretary of the Department of Health and Human Services (DHHS) evaluate and address programs that aim to reduce disparities in screening, early detection, and treatment of cancer among ethnic and racial minorities, specifically among AI/AN, A/PI, African-Americans, and Hispanics. CMS has awarded cooperative agreements to six sites: Johns Hopkins University, the Josephine Ford Cancer Center, the MD Anderson Cancer Center, the Huntsman Cancer Institute, Molokai General Hospital, and the New Jersey Medical School. Each initiative is described below in Section 1.3.

Each demonstration project will have both a screening arm and a treatment arm. Eligibility is limited to the minority group selected by the demonstration; all participants must be enrolled in both Parts A and B fee-for-service Medicare and, according to the RFP, not reside in institutions. Upon intake, participants will complete a Cancer Screening Assessment (CSA), which will be used (among other things) to assign participants to the screening arm (if they do not have cancer) or to the treatment arm (for those diagnosed with cancer). CMS' implementation contractor will randomize participants within each arm to either the intervention group or the control group. The screening intervention group will receive facilitation services designed to improve cancer screening rates in accordance with screening guidelines and to ensure that participants who screen positive receive the necessary diagnostic work-up to confirm or rule out a cancer diagnosis. The screening control group will receive "usual care," i.e., whatever screening or diagnostic services they seek on their own and/or are recommended by their providers. Participants in the treatment arm will all have a cancer diagnosis and may or may not have already started treatment. The treatment intervention group will receive facilitation services designed to ensure completion of all primary and secondary cancer treatments, as well as all necessary follow-up and monitoring. The treatment control group will receive "usual care," i.e., those cancer treatment services recommended by their providers that they choose to receive.

1.3   Summary of the Demonstration Sites

Table 1 summarizes each site's proposed approach. More detail is provided in the text below.

Huntsman Cancer Institute. The Huntsman Cancer Institute will utilize patient navigators in the treatment group of the screening and treatment arms with the focus on rural American Indians in the Intermountain West (Utah and Montana). Navigators, who will be tribal members, will be trained to guide peers through the process of cancer screening, diagnosis and treatment. However, instead of randomizing individuals into the treatment and control groups, the Institute will randomize tribal locations. This is to prevent contamination of the groups in the small tribal communities.

The Institute plans to recruit 1,800 men and women for the intervention and control groups of the screening arm, and 140 for the intervention and control groups of the treatment arm. Each will receive a $10 incentive for completing a CSA.

Johns Hopkins University. Johns Hopkins University (JHU) plans to compare the efficacy of a nurse-led team of community health workers to that of a less intensive intervention in facilitating adherence to recommended screenings and treatments among elderly African Americans in Baltimore City, Maryland. The less intensive intervention consists of usual medical care for cancer screening, treatment and diagnosis and a feedback report to the beneficiary that includes general information about cancer and Medicare-covered services. The more intensive intervention consists of community health workers trained to provide education and counseling to beneficiaries to encourage cancer screening and treatment. The counseling will be conducted in a culturally appropriate manner and will address the beneficiary's knowledge of cancer as well as any barriers that the beneficiary has toward seeking care. The community health worker will also assist the beneficiary in scheduling appointments and keeping them, including accompanying the beneficiary to the appointment, if necessary.

Table 1
Overview of CPTD Sites

Grantee	Location of Demonstration	Type of Program	Priority Population	Rural / Urban	Sample Sizes	Notes
Huntsman Cancer Institute at the University of Utah, Salt Lake City, Utah	Tribal locations in the Intermountain West (Utah and Montana)	Intervention=patient navigation Control=usual care	American Indians	Rural	Screening=1,800 Treatment=140	randomization of tribal locations $10 for completing CSA
Johns Hopkins University, Baltimore, Maryland	Baltimore City, Maryland	Intervention=nurse-led team of community health workers Control=usual care and feedback report	African Americans	Urban	Screening=2,874 Treatment=400	$10 incentive to complete CSA at entrance, $5 at exit from demonstration
Josephine Ford Cancer Center-Henry Ford Health System, Detroit, Michigan	Detroit, Michigan (Wayne County)	Intervention = nurse/patient advocate partnership Control = usual care, reassessment at end	African Americans	Urban	Screening=1,900 Treatment=1,150	$25 incentive to complete CSA financial assistance for copayments and coinsurance
MD Anderson Center's Center for Research on Minority Health and CHRISTUS St. Joseph Hospital, Houston, Texas	Houston, Texas (Harris County)	Intervention=patient navigation Control=usual care	Hispanic/ Mexican Americans	Urban	Screening=3,240 Treatment=360	
Molokai General Hospital, Molokai, Hawaii (Maui County)	Molokai Island and Oahu, Hawaii	Intervention = patient navigation on Molokai and / or on Oahu Control = usual care and Nutrition Education Program	A/PIs	Rural	Screening = 528 Treatment = 50	
New Jersey Medical School, Newark, New Jersey	Newark, New Jersey	Intervention=community health worker Control=usual care	Hispanics	Urban	Screening = 1,284 Treatment=100	$20 prepaid phone call to complete CSA at entrance, $20 prepaid phone call at exit from demonstration

JHU plans to recruit 2,874 men and women for the intervention and for the control groups of the screening arm, and 400 for the intervention and for the control groups of the treatment arm. JHU will give participants a $10 incentive for completing the CSA and $5 at exit of the demonstration project.

Josephine Ford Cancer Center. The Josephine Ford Cancer Center plans to utilize a nurse/patient advocate partnership model for elderly African Americans in Detroit, Michigan. Each partnership will have its own panel of participants to follow. The nurse will delegate appropriate tasks to the patient advocate and respond to medical questions and issues. Participants assigned to the control group will receive usual care and be reassessed at the end of the project.

The Center plans to recruit 1,900 men and women for the intervention and control groups of the screening arm, and 1,150 for the intervention and control groups of the treatment arm. As an incentive the Center will provide all beneficiaries who had not previously been diagnosed with cancer a $25 grocery store gift card. There will also be financial assistance for copayments and coinsurance for the screening and treatment groups.

MD Anderson Cancer Center. The MD Anderson Cancer Center will utilize patient navigators to assist patients from point of recruitment, assessment and screening to care and appropriate monitoring and follow-up. Both professional (physicians, nurses, social workers) and lay navigators will serve as patient navigators of Medicare beneficiaries of Hispanic ethnicity, specifically of Mexican American origin, living in Houston, Texas.

Beneficiaries without a previous diagnosis of cancer will be placed in the screening arm and randomized into the control or intervention group. The control group will receive their usual source of care. The intervention group will be assigned a patient navigator to ensure appropriate screening and follow-up. Beneficiaries with a diagnosis of cancer will receive a patient navigator to assist them in receiving appropriate treatment.

The Center has a goal of having 3,600 unique beneficiaries completing exit interviews, 3,240 men and women for the intervention and control groups of the screening arm, and 360 for the intervention and control groups of the treatment arm.

Molokai General Hospital (Kukui ahi Program). This Program will take place on the island of Molokai, which is one of Hawaii's smallest inhabited islands with approximately 7,400 residents. Molokai is a Medically Underserved Area and a Health Professional Shortage Area for Primary Care, Dental Care and Mental Health. Although some screening and diagnostic tests can be done on the island, many beneficiaries need to travel to Oahu for them. Most cancer patients need to visit Oahu at least once during their treatment.

Molokai General Hospital will institute two forms of patient navigation for the Kukui ahi Program. The grantee was concerned that it would appear to island residents that only the intervention group received services and the control group did not, and this would affect beneficiary participation. To counteract that perception, beneficiaries in the control groups of the treatment and screening arms will receive their usual source of care in addition to a Nutrition

Education Program (NEP). The NEP will consist of cooking demonstrations, meal planning and education on nutrition and health, focusing on those with diabetes, hypertension and arthritis.

The intervention is community navigation. On the island of Molokai navigators will provide culturally appropriate materials and assistance with making appointments and transportation for the screening arm. The intervention group of the treatment arm will consist of beneficiaries already diagnosed with cancer. They will receive a professional navigator on the island of Oahu for their visits for diagnostic tests or oncologist visits. The control group of the treatment arm will be identified from a section of Maui called Hana. The population in Hana is nearly identical in make-up to that of Molokai.

The focus of the demonstration is A/PI. The Program plans to recruit 528 men and women for the intervention and control groups of the screening arm, and 50 for the intervention and control groups of the treatment arm.

New Jersey Medical School. The New Jersey Medical School plans to implement a community educational and outreach program along with a community and hospital navigator program. These programs will help beneficiaries navigate the health care system, help facilitate adequate screening, diagnosis and treatment of cancers, and provide patient assistance in all aspects of the health care continuum. The Demonstration will focus specifically on Latino Medicare beneficiaries, the largest percentage of whom are from Puerto Rico and Ecuador, in Newark, NJ.

The School plans to recruit 1,284 men and women for the intervention and control groups of the screening arm, and 100 for the intervention and control groups of the treatment arm. Each participant will receive a $20 prepaid phone call after completion of the initial CSA and after completion of the final CSA.

1.4   Overview of Report

The remainder of the report is organized as follows: Section 2 provides an overview of the evaluation, including research questions and methodological challenges. Section 3 discusses the site visits and focus groups. Section 4 describes the beneficiary survey and analysis. Section 5 reviews the Medicare claims analyses. Section 6 describes the cost-effectiveness and cost utility analyses. Section 7 outlines the reports that will be produced by the evaluation; and Section 8 contains a timeline and deliverable schedule.

SECTION 2
EVALUATION OVERVIEW

2.1   Research Questions

Key research questions are described in ***Table 2***, along with our proposed technical approach and data sources. This list is not meant to be complete or exhaustive, rather it summarizes our basic approach to the principal research questions. They encompass the questions specified in the RFP, as well as additional questions raised by CMS (e.g., baseline disparities).

2.2   Design Overview and Challenges

Our proposed evaluation design is intended to integrate multiple analyses, involving both quantitative and qualitative data, to meet the requirements of the Reports to Congress, including: (1) whether racial/ethnic disparities were reduced by improving screening rates, treatment completion rates, etc.; (2) whether Medicare spending was reduced (or at least not increased); and (3) whether beneficiaries and providers were satisfied with the demonstration services. The randomized nature of the design is a major strength, as we can rule out selection bias and other factors from any observed differences. We will conduct two evaluations at each site, one for each arm of the demonstration: (1) screening intervention vs. control group beneficiaries, and (2) treatment intervention vs. control group beneficiaries. Medicare claims, merged with project demonstration data, will be used to compare screening and diagnosis outcomes (or treatment outcomes, depending on the study arm), as well as Medicare costs and service utilization. Claims-based analyses will include all demonstration participants. A subset of participants in both arms (intervention and control) will be surveyed; survey analyses will compare satisfaction, quality of life, perceived barriers, and utilization of non-Medicare services. We will merge survey responses with claims data and conduct cost-effectiveness and cost-utility assessments of each demonstration. We will also collect and analyze program cost data at each site. Findings from site visits will be used to explain any differential outcomes within and across demonstration sites.

Our proposed design is described in detail in the following chapters. But first we discuss some potential evaluation challenges and how we plan to address them.

Contamination of control group. Because of the nature of these demonstrations, there may be spillover effects from the intervention on control group members. Control group members may learn about the importance of cancer screening by word of mouth from intervention group members in their community, or community health workers may be unable or unwilling to exclusively target their efforts to intervention group members. We propose two strategies to measure the possible contamination in the control group (described in more detail in later sections). First, we will construct a comparison group from the Medicare EDB that "looks like" the demonstration population with regard to race/ethnicity and geographic location and compare their utilization with that of the intervention and control groups in the screening arm. Second, our beneficiary survey will ask both intervention and control groups about their use of facilitation services. This will allow us to determine whether control group members received

Table 2
Overview of the CPTD Evaluation

Research Question	Technical Approach	Data Sources
Impacts on Cancer Screening and Treatment		
Did the screening intervention increase screening rates? Did it improve adherence to cancer screening guidelines?	Calculate and compare percent screened for intervention and control groups.	Physician and HOPD Part B claims data Cancer screening guidelines
Did the screening intervention increase the percent of beneficiaries with abnormal screening results who reach a definitive diagnosis? Was the time from screening to definitive diagnosis shorter? Did it improve adherence to diagnostic guidelines?	Calculate and compare the percent of intervention and control groups who receive appropriate diagnostic follow-up. Calculate and compare the time intervals between tests.	Physician and HOPD Part B claims data Demonstration site data on positive screens (if available) Cancer practice guidelines
Did the screening intervention increase the percent of cancers that were diagnosed early?	Calculate the percent of definitively diagnosed cancers that were diagnosed at an early stage (before metastasizing). Compare screening and control groups.	Demonstration site on positive screens and stage at diagnosis (if available)
Did the treatment intervention increase completion rates for primary treatment?	Calculate and compare the percent of treatment and control group who receive surgery.	Physician/supplier and HOPD claims data Demonstration site data on cancer stage at diagnosis (if available)
Did the treatment intervention increase completion rates for multi-modality treatment? Did it shorten the time interval from start to finish of treatment? Did it improve adherence to treatment guidelines?	Calculate and compare the percent of treatment and control groups who receive surgery plus subsequent chemotherapy and/or radiation therapy	Physician/supplier and HOPD claims data Demonstration site data on cancer stage at diagnosis (if available) Cancer practice guidelines
Did the treatment intervention increase post-treatment follow-up rates?	Calculate and compare the percent of treatment and control groups who make follow-up visits to oncologists and/or primary care physicians.	Physician claims data
Impact on Medicare Spending		
Did the demonstration increase Medicare spending for intervention participants? If so, what was the source of the increase (type of service)?	Calculate total Medicare expenditures per beneficiary (in total and by type of service). Compare screening intervention vs. control group and treatment intervention vs. control group.	Medicare physician/supplier, hospital, HOPD, and home health claims data
Did the intervention increase utilization of non-cancer related Medicare services? If so, what types of services contributed to the increase?	Calculate utilization rates and Medicare expenditures per beneficiary for cancer-related vs. non-cancer-related services. Compare screening intervention vs. control group and treatment intervention vs. control group.	Medicare physician/supplier, hospital, HOPD, and home health claims data
What is the cost-effectiveness of the screening and treatment intervention groups relative to their control groups? What is the cost-utility of the screening intervention? Of the treatment intervention?	Construct cost-effectiveness ratios (cost per person screened, cost per patient receiving appropriate treatment, etc.) Construct cost-utility measures (cost per QALY). Compare screening intervention vs. control group and treatment intervention vs. control group.	Medicare physician and HOPD claims Capitation payments for facilitation services Demonstration program data Beneficiary survey

(continued)

Table 2
Overview of the CPTD Evaluation (continued)

Research Question	Technical Approach	Data Sources
Impact on Beneficiary Outcomes and Satisfaction		
Did the screening intervention improve quality of life for participants? Did the treatment intervention?	Calculate quality of life measures (e.g., SF-12, FACT-G, EQ-5D) for each beneficiary). Compare screening intervention vs. control group and treatment intervention vs. control group	Baseline CSAs Beneficiary survey
Did the screening intervention improve satisfaction for participants? Did the treatment intervention?	Calculate and compare satisfaction measures for intervention and control groups (screening and treatment separately).	Beneficiary survey
What factors appear to improve quality of life and satisfaction?	Analysis of barriers affecting access and utilization of non-Medicare services. Comparison of intervention and control groups.	Beneficiary survey
Provider Satisfaction		
Are providers satisfied with the facilitation services offered by the demonstration?	Structured interviews with providers.	Site visits
Do they believe these services are improving quality of care for those in the intervention group?	Structured interviews with providers.	Site visits
Do providers feel that the demonstration has improved outcomes for those assigned to the Intervention group?	Structured interviews with providers.	Site visits
Scope of Facilitation Services		
What screening and treatment facilitation services are provided? Which ones have been most/least effective, and why?	Structured interviews with demonstration site staff.	Site visits Written materials provided by demonstration sites
How have participants responded to the services that are offered? Are there services that have been refused?	Structured interviews with demonstration site staff.	Site visits
Success of Demonstration in Meeting Its Goals		
Has the demonstration project been successful in reaching its target population?	Compare number of participants recruited with number estimated in site's application.	Secondary data from demonstration site
How has the project sought to prevent contamination of the control group? Has this been successful?	Review site's plans for avoiding contamination. Structured interviews with project's clinical and community staff. Comparison of receipt of facilitated services between intervention and control groups.	Secondary data from demonstration site Site visit Beneficiary survey
Has the demonstration succeeded in organizing and delivering facilitation services?	Structured interviews with site staff. Survey data for intervention groups.	Site visits Annual CSAs
How has the demonstration project sought to ensure improvement in cultural competency?	Structured interviews with demonstration site staff.	Site visits
Has the demonstration project made any changes to the health care system (in addition to the facilitation services) to improve care?	Structured interviews with demonstration site staff.	Site visits

(continued)

Table 2
Overview of the CPTD Evaluation (continued)

Research Question	Technical Approach	Data Sources
Baseline Disparities		
What are the current disparities in cancer screening between racial/ethnic minorities and white beneficiaries? Do those disparities differ across geographic areas?	Calculate and compare percent screened for minority groups vs. whites.	Physician and HOPD Part B claims Cancer screening guidelines
What are the current disparities in cancer treatment between racial/ethnic minorities and white beneficiaries? Do those disparities differ across geographic areas?	Calculate and compare the percents of minority groups vs. whites who receive surgery, who receive subsequent chemotherapy and/or radiation treatment, and who receive follow-up visits.	SEER-Medicare claims data Cancer treatment guidelines

such services, whether from demonstration staff or from someone else (e.g., providers elsewhere, family, friends).

Deviations from randomization. Two of the demonstration sites have rejected randomization as originally proposed by CMS, i.e., to randomize by individual participant. One site has proposed to randomize communities, instead of individuals. The second site proposes a nonequivalent comparison group. Both deviations obviously weaken the experimental design, although they may also reduce contamination. Because each of these sites is unique in terms of the minority group it serves, findings regarding the impact of facilitation services on American Indian and Native Hawaiian beneficiaries may be inconclusive.

Short-run increases in Medicare spending by intervention groups. It is possible that Medicare spending (over and above the facilitation payments) will actually increase in the intervention group relative to the control group, especially given the relatively short duration of the demonstration. (Earlier detection of cancer should reduce cancer treatment costs in the long-run, but any such cost-savings may not be observed during the time frame of this evaluation.) Spending may increase for several reasons: (1) if intervention group members receive screening or other cancer-related services that they otherwise would not have gotten; (2) if intervention group members receive cancer treatment earlier than they otherwise would have (and hence those costs are captured during the time frame of this evaluation); and/or (3) if the interventions raise health consciousness and lead intervention group members to seek more non-cancer-related services. Spending comparisons over the short-term are inappropriate, given the potential for reduced downstream spending and the added gain in quality-adjusted life years (QALYs). We propose an additional, more appropriate, comparison: differences in cost per QALY and cost per unit of improvement between intervention and control groups (using cost-utility and cost-effectiveness analysis, respectively).

Potentially low enrollment in some demonstration projects. It is likely that some projects will be unable to enroll sufficient participants to support the quantitative analyses we propose. In these instances, we propose that the evaluation of such projects be limited to site visits and claims analyses. We also propose that the potential survey samples for these projects

be reallocated to those that did achieve sufficient enrollments. This would increase the survey sample sizes for those sites and increase our power to detect difference in outcomes.

Differential screening success by type of cancer. It is possible that demonstration projects will show mixed success for the screening intervention, achieving better outcomes for some cancers compared with others. It may be easier, for example, to increase screening rates for those tests that are less invasive (e.g. PSA tests) as opposed to flexible sigmoidoscopy or colonoscopy. We will examine screening outcomes separately by type of cancer (targeted by the test) using claims data. We will use site visit interviews and annual CSAs to determine whether facilitation services were more or less successful with the different cancer screening tests.

SECTION 3
SITE VISITS

Combining quantitative and qualitative methods is a common approach to strengthen a study (Patton, 2002). This technique is referred to as triangulation. Our experience with qualitative methods and analytical techniques (as described in Kreuger & Casey, 2000; Miles and Huberman, 1994) will provide CPTD with useful formative feedback, along with qualitative summative data, for addressing complicated multilevel and cross-program assessments related to provider satisfaction, scope of facilitation services, and success of demonstration in meeting its goals. This Section includes a brief overview of the qualitative methods and process that will be used for the CPTD evaluation.

3.1   Purpose of Site Visits

Two sets of site visits are planned for the CPTD evaluation. The first set will occur within six months of program initiation, while the second will occur during the third year (or toward the end) of funding. All sites will be visited during each of these sets of site visits. The following provides an overview of the purposes for each set of site visits.

3.1.1. Initial Site Visit Purposes

An initial site visit to each of the six grantee sites will take place in the summer to early fall of 2006, or within six months of program initiation. The overall purpose of the initial site visits is to gain a thorough understanding of each grantee's processes and program, as well as to address data collection issues that each site may encounter in providing required evaluation data. This information will be used to:

- Develop an in-depth understanding of the interventions each grantee is implementing and the processes they are following to recruit patients and provide services;

- Identify barriers and issues that are common across the sites;

- Identify technical assistance needs;

- Obtain examples of exemplar strategies or processes being used by grantees; and,

- Resolve issues that may impact the evaluation of the project such as in administering the study instruments, etc.

In order to begin to understand the sites, RTI has reviewed the grant applications to identify commonalities across the sites, unique characteristics of each and to begin to characterize the sites themselves. For example, by extracting information about the lead agency at each site, the priority population, details on the interventions to be implemented, and other information, RTI is already somewhat familiar with the sites and will be able to identify initial strengths and weaknesses to program implementation. RTI has created a summary matrix of the sites in order to begin to describe what they are doing, the populations they are targeting, as well as other issues (*Table 1*). More information about each site will be obtained at the May 2006 meeting of all demonstration projects. Details obtained during the Grantee meeting in May will be

incorporated into the summary matrix so that when the site visits are conducted, RTI will have the most up-to-date descriptions of each site. Prior to each of the initial site visits, RTI will use the information from the applications as well as other available details to gain an in-depth understanding of what that site plans to implement and how.

3.1.2. Follow-up Site Visit Purposes

A second set, or follow-up, site visits to all of the demonstration programs are scheduled to take place in Spring 2009. The purposes of these subsequent site visits will be to gain a thorough understanding of how they have implemented their programs, what challenges they have encountered and what successful resolutions they found, perceptions of how successful the project has been, among others. During these visits, we will also be interested to learn more about how they collaborated with local partners, the most successful strategies for doing so, and the organizational and/or system changes that have occurred as a result of these projects. In preparing for each type of site visit, RTI will need to develop thorough protocols as guides for conducting interviews and plan and implement the visits as follows.

3.2   Site Visit Plan

3.2.1 Site Visit Protocol Development

For the initial site visits, since the programs will be relatively 'young' in their stage of development, there will likely only be a few staff and possibly providers whom we can interview. After (or during if possible) the Grantee meeting in May, RTI will conduct an initial telephone interview with the PI for the selected site in order to identify the people who could be interviewed during the site visit and obtain an update on their project implementation. RTI will then draft protocols for the people we plan to interview. Issues to address in the initial site visit are likely to include:

- Thorough description of their program, including what they plan to do and who will be involved

- Description of how they plan to recruit participants into each arm of the study

- Barriers they anticipate in recruiting participants and implementing the study

- Anticipated problems with administering demonstration data collection and transmission of data to RTI.

Other issues may be identified to discuss during this site visit as we learn more about each site and the interventions they plan to implement. Once these issues are identified, RTI will draft protocols for the initial site visit for review and comment by CMS. Once finalized, we will obtain RTI IRB approval for the protocols.

The same process will be followed in preparing for the site visits that occur later in the project. Because it is important to capture a comprehensive view of the program, the interviewees will represent multiple levels of the program. For these site visits, we anticipate meeting with the PI of each site, support staff, perhaps key providers, and others as identified by

the site in order to ask them detailed questions about their program and the process of implementation. Protocols for each type of respondent will be drafted based on the evaluation questions for the study (see **Table 2**) for many of these questions). When possible, we will incorporate similar items across the protocols in order to identify common themes from the data during analysis. The following describes the process for conducting all of the site visits.

3.2.2 Conducting the Initial and Follow-up Site Visits

The process for conducting both the initial (Year 1) and follow-up (Year 3) site visits will be similar. RTI will conduct semi-structured interviews with the staff at each site. When feasible, we will also observe clinic operations and patient interactions, and obtain documents that provide details about the program. Prior to each planned visit, RTI will contact the PI for the project and obtain their recommendations of who to interview during the site visit. The number of people to be interviewed at each site will vary, based on the size of the program, the number of partners and/or providers who have been involved, and the geographic distance from facilities working on the project. For example, clinics working on the Huntsman Cancer Center project have been described as being several hundred miles apart. Due to the limited time at each site and the infeasibility of visiting all of the clinics, it is likely that only 2-3 clinics working with this project will be able to be visited. Attempts will be made to visit different clinics during the initial site visit than those visited in the follow-up site visit if that seems to be an important assessment to make in the evaluation.

For the initial site visits, we anticipate that only about 5-8 people will be able to be interviewed. Since the programs will be early in their development, such that partnerships have yet to be formally established, all staff have not been hired, among other issues, there will likely be fewer people knowledgeable enough about the program to be interviewed. For this reason, we estimate that each of the initial site visits will require 1-2 days. The conduct of the initial site visits include a slightly different team from RTI than for those conducted during Year 3. While a primary purpose of the initial site visits is to collect data about how the grantees are operating, another purpose is to ensure that the sites understand the data collection requirements for the CPTD evaluation. Therefore, for the sites in Texas, New Jersey, Maryland, and Michigan, two RTI staff will conduct the site visits for the purposes of data collection, while a third staff member will participate in the site visit to provide training and technical assistance on the evaluation measures and instruments. For the remaining sites in the far Western United States and Hawaii, because of the high cost of travel, two RTI staff will conduct these site visits and will collect the data as well as provide technical assistance on ongoing data collection.

Depending on the size of the staff and the number of people who have been involved in implementation of the program, we anticipate far more people will be interviewed from each site during the follow-up site visits. We estimate between 10-15 interviews to be conducted during each of the follow-up site visits. Depending on the scheduling of these visits and allowing for time to travel, we have planned for a 2-3 day trip to each of the sites, with at least two RTI staff traveling to each site.

For each set of site visits, RTI will contact each site prior to the visit in order to coordinate scheduling of interviews and obtain lists of individuals we should plan to meet with. We will ask each PI to identify a contact person for us who can help facilitate our visit. During

each individual or group interview, one RTI staff member will facilitate the discussion while another serves as note taker. The note taker will ensure that, with respondents' permission, the interviews are digitally recorded and detailed notes are obtained. After each interview, the team will conduct debriefings to record any observations or impressions. During all of the site visits, we will also gather any program documents that will help to describe the programs. Upon return from each site visit, the notes will be typed and entered into a qualitative software program, *Atlas.ti*, so that coding of the data can begin.

3.3   Analysis Plan for Site Visit Data

As described, each interview, whether of a group (if that's the most feasible way of obtaining data) or with an individual, will be digitally recorded. Data collected during the initial site visit will not be analyzed to the extent that we will analyze data collected under Task 10. The analysis for the initial site visit will include a summary description of each of the sites.

For the visits under Task 10, the audio tapes and notes from all the discussions will be combined and edited for accuracy in order to be exported into the qualitative analysis software, *Atlas.ti* (version 5.0) to prepare for coding of the data. The *Atlas.ti* software offers tools to manage, extract, compare, explore, and reassemble large amounts of data in a systematic way. Its focus is on the analysis of qualitative, rather than quantitative, data. Qualitative data are imported into *Atlas.ti* and assigned as a "primary document" or PD. PDs represent the textual, graphical, audio and/or video data the researcher wishes to interpret (i.e., the data source). Multiple and varying types of PDs can be saved as a single *Atlas.ti* project. For example, all transcribed data from each of the grantees will be saved as a single project file. Having a single data file facilitates cross-comparison of data and co-authoring in that it allows two or more researchers to work on (code) the same project, while keeping the respective sources of ideas identifiable at all times.

Many believe that a problem with qualitative research is that it is done chiefly with words, not with numbers, thereby lending itself to bias in its interpretation. However, there are many strategies that can be used to increase the reliability of findings and eliminate the subjectivity of interpreting results. A common solution to this issue is to code the transcriptions, field notes, and observations, using a consistent set of terms. A code is "an abbreviation or symbol applied to a segment of words—most often a sentence or paragraph of transcribed field notes—in order to classify the words" (Miles & Huberman, 1994, p. 56). Codes are really categories that are typically derived from the research questions, key concepts, or important themes. To prepare for coding the data for the CPTD, the following methical steps will be taken:

1. Creation of codes
   The most effective way to create codes is to start prior to the fieldwork, or site visits in the case of the CPTD, by listing common characteristics or themes that are expected, based on the study's research questions. This list of codes is ideally derived from the study's conceptual framework, research questions, hypotheses, and/or key variables that the researchers define during the design of the study. There are three primary types of codes that are generally used, varying by the degree of inference used in making them. These codes include those that are *descriptive*, in that they

entail no interpretation but "simply attribution of a class or phenomena to a segment of text"; *interpretive* where text is coded based on the researcher's interpretation of the local dynamics, or context to indicate the meaning of the text, and finally, even more inferential or *explanatory* codes that indicate text that support emerging themes or leitmotiv that the analyst has "deciphered while unraveling the meaning of local events and relationships" (Miles & Huberman, 1994, p. 56). While developing the site visit protocols, RTI will identify a number of key factors that could be impacting how well the grantees are able to implement their programs and use these as our initial codes.

Upon return from the site visits, we will then convene our group of researchers to debrief and identify emerging themes from our initial impressions from our trips. When one is working with text, or less well-organized displays, it is often possible to "note recurring patterns, themes, or 'Gestalts', which pull together a lot of separate pieces of data" (Miles & Huberman, 1994, p. 216). It is often useful to discuss general impressions across sites in order to determine if there are emerging themes that can be identified. Through this discussion, codes can be refined and preparations made to begin the coding process.

2. Defining codes
   A key step in preparing for data analysis is in systematically defining each code so that independent analysts are more likely to code an array of text with the same code. This step enhances rater-reliability and ultimately improves the quality of the interpretation of the data. As noted by Miles and Huberman, "clear operational definitions are indispensable, so that [codes] can be consistently applied by a single researcher over time, and so that multiple researchers will be thinking about the same phenomena as they code" (1994, p. 60). Initially, each code will be defined by a small group of researchers for the CPTD who will then discuss these with the larger group of analysts. Each code will be thoroughly discussed so that there was consensus over the meaning of each.

3. Testing and refining codes
   During the first stages of coding the data, it is important to test the codes and then reconvene the researchers so that they can discuss their assumptions and ensure that they are consistently coding similar text the same way. At this stage, some codes are likely to change. Through the initial testing of the codes, it is likely that some will be identified as inappropriate or inadequate in describing particular phenomena that the text is conveying. In addition, other codes may begin to emerge as more in-depth review of the data begins and themes become more apparent.

To further increase reliability, data from each grantee will be coded by two researchers who work independently and concurrently to code their data set. These assignments were made in order to double code the data. As noted by Miles and Huberman (1994, p. 60), "definitions get sharper when two researchers code the same data set and discuss their initial difficulties". Double-coding not only aids definition clarity, but is a good reliability check. Since two different people are coding the data, they have the opportunity to check their impressions and interpretations with each other, thereby increasing the extent to which the data are reliably

analyzed. Since site visits will be involved, it's also important to include in these teams one person who was present at the site, and therefore has already formed some initial impressions about its findings but also has knowledge of the context, and another who was not present so that they are completely unbiased in terms of what was seen and heard.

It is anticipated for the CPTD that the priority findings from the follow-up site visits are those 'across site' or that seem to be common themes across the sites, as opposed to a 'within site' analysis that would provide specific findings for each grantee. RTI will analyze the findings in order to provide CMS with a summary report of lessons learned, impressions, recommendations, among others, that can be incorporated into the report to Congress and used to explain some of the quantitative findings.

SECTION 4
BENEFICIARY SURVEY

4.1   Survey Objectives

Separate questionnaires will be developed for beneficiaries in the treatment and screening study arms. These questionnaires are designed to monitor important outcomes that are not captured by claims data. The primary outcomes of interest are beneficiary quality of life, satisfaction with care, and health utilities. The questionnaires will also collect self-reported health status that will be used in the cost-effectiveness analyses. We have selected brief, standardized scales with demonstrated reliability and validity in older populations to measure these outcomes. These scales are also general enough to apply to both the intervention and control groups.

4.2   Proposed Survey Measures

Given the differences between the two study arms, we recommend that generic measures be used for the screening arm and that measures for the treatment arm be tailored to cancer care. *Table 3* presents a preliminary list of potential scales for the surveys. All of the major outcomes in the table have also been incorporated in the CSA so that they will be collected at baseline as well as the survey follow-up. The EQ-5D is to be included in the both arms to collect utility data needed for the cost-effectiveness analysis.

Table 3
Proposed Beneficiary Survey Measures

Domain	Screening Arm Survey	Treatment Arm Survey
Satisfaction	Consumer Assessment of Healthcare Providers and Systems (CAHPS; scales for access and interaction with healthcare providers)	EORTC Comprehensive Assessment of Satisfaction with Cancer Care
Quality of Life	VR-12/SF-12 (physical and mental functioning)	Functional Assessment of Cancer Therapy-General (FACT-G; physical, social/family, and emotional well-being)
Quality-Adjusted Life Years	EQ-5D (utility)	EQ-5D (utility)
Use of other services	Use of other providers to obtain screening or similar facilitation services	Use of other providers for cancer care and facilitation services
Reasons for non-compliance	Barriers to screening	Barriers to using facilitation services; use of other treatment modalities

Screening Arm Survey. A variety of survey instruments are available to measure overall functioning and quality of life in the general adult population. Of these, the SF-36 has been used most frequently with Medicare beneficiaries. An even briefer version of the original instrument, the SF-12, will provide valid estimates of physical and mental health. One advantage of using the SF-12 is that benchmark values are available for purposes of comparison. Some recent developments with the SF-12 allow values for mortality to be imputed so that deaths can be included in analyses. We will use a version of this instrument known as the VR-12 that is now being fielded in several large CMS studies. The most widely used satisfaction measure for Medicare beneficiaries is the Consumer Assessment of Healthcare Providers and Systems (CAHPS) instrument. From the latest version of this instrument, we are recommending 8 items. These items measure how much of a problem it has been to get care and the quality of interpersonal interactions with health care providers.

Treatment Arm Survey. While the SF-12 and the selected CAHPS could also be used in the treatment arm questionnaire, more sensitive analyses of the treatment facilitation services can be achieved by employing items from cancer-specific survey instruments. A widely used instrument in cancer studies is the Functional Assessment of Cancer Therapy-General (FACT-G) which addresses issues applicable to patients living with a variety of different types of cancer. Version 4 of the FACT-G consists of 27 items measuring four domains (physical, emotional, social/family, and emotional well-being). The reliability, validity, and responsiveness of FACT-G scales have been supported by a number of studies.

In the treatment arm, the satisfaction measure should be tailored to cancer care. It should be broad enough so that it is relevant to the usual types of care received by the control group, but sensitive enough to be responsive to the types of additional facilitation services that will be provided to intervention patients. The recommended satisfaction measure is the European Organization for Research and Treatment of Cancer Comprehensive Assessment of Satisfaction with Care Questionnaire (EORTC QLQ- SAT32). We selected 12 items from this instrument that evaluate the technical skills, interpersonal skills, information provision, and availability of cancer care providers.

Health Utilities. As discussed in Section 6, a serious shortcoming in the field of cancer research is the lack of data on utilities (preferences for health states measured on a continuum ranging from 0=death to 1=optimal health) that can be used for cost-utility analyses. Utilities be measured directly in the survey groups using the EQ-5D. This very brief 5-item measure, consisting of questions about mobility, self-care, usual activities, pain, and anxiety/depression, produces utility scores for individual respondents. Weights reflecting societal preferences in the United States have recently been published on the basis of a large, representative national survey (Shaw et al., 2005).

Survey Pretesting. We will pretest each survey instrument with 9 eligible beneficiaries. Two sites, one with English-speaking and one with Spanish-speaking beneficiaries, will participate in the pre-testing phase. In each site, 4 beneficiaries will be assigned to take the Screening survey and 5 cancer survivors will be asked to take the Treatment survey.

We will use cognitive interviewing techniques to pretest the items in the mail surveys. While cognitive interviews are frequently conducted in person, we recommend testing the

beneficiary instrument for this study using a combination of a mail survey and structured telephone interviews. Telephone interviews are less costly and time-consuming (due to the lack of travel), and often more convenient for patients. Testing via structured telephone interviews is likely to provide the most information with the lowest respondent burden. We have used this approach successfully on a number of current and previous studies for CMS, including the *Medicare CAHPS Disenrollment Survey* and the *Medicare CAHPS Fee for Service Survey.* Pretest respondents will receive financial incentives, such as a $20 gift card.

RTI survey methodologists and specialists will test the questionnaires by mailing participants a copy of the questionnaire and then calling soon after to collect the participants' responses and debrief on selected questionnaire items. We will prepare debriefing notes after each interview describing any problems detected with the survey questions. Once all interviews are completed, our methodologists will meet to identify questions that were problematic and suggest possible revisions to improve the questionnaires. We will submit a memorandum summarizing the testing results and recommend changes to the wording, structure, and content of the final questionnaires.

4.3   Sample Size and Power

The survey design calls for a total of 5,000 beneficiary surveys, 3,000 in the screening arm and 2,000 in the treatment arm. Because the length of the recruitment period varies by site, we will allocate these surveys among beneficiaries who completed baseline CSAs at least 6 months earlier. Analyses will be performed separately for each site.

Screening Arm. The target sample size in this arm is 500 completed questionnaires per site, consisting of 250 beneficiaries from the intervention group and 250 from the control group. Survey respondents will be randomly sampled from those completing the baseline CSA.

One important feature of the design is that baseline measures for each survey outcome will be available from the CSA. Controlling for baseline status will help to increase the power of statistical tests. Using conventional assumptions for statistical power (power=.80, one-sided tests, and alpha=.01 to compensate for multiple outcomes) and assuming that covariates explain 25% of the variation in an outcome, a sample size of 250 per group will permit us to detect an effect size (mean group difference divided by the standard deviation of the outcome measure) of .26 or more for each site. This is close to what are generally regarded to be a "small" effect, and is equivalent to an absolute difference of 2.6 points between groups on the SF-12 component scores.

Treatment Arm. The number of beneficiaries to be recruited for the treatment arm will be much smaller than in the screening arm. The estimates reported by the sites range from only 50 cases to 400 cases, with a sixth site listing 1,150 cases. Many of these cases will not be identified until after the survey is scheduled to begin. As a result, we will survey all participants in this arm of the study. The small sample sizes in some sites mean that only comparatively large effects can be detected at the site level.

4.4   Data Collection

The beneficiary survey will be conducted by mail with telephone follow-up of nonrespondents. Our protocol is based on the Dillman Total Survey Design method that RTI has successfully implemented on many previous Medicare studies. We will begin by sending a prenotification letter, followed by the first questionnaire mailing. Approximately one week later, a thank you/reminder postcard will be sent. We will send a second questionnaire package to all sample members who do not respond to the first questionnaire, approximately two weeks after the thank you/reminder is mailed. We will follow up by telephone for all nonrespondnets for whom we can obtain a telephone number. We will send a third questionnaire package via overnight mail to all nonrespondents for whom we are unable to obtain a telephone number.

We will work closely with CMS to prepare all mail survey cover letters and other materials, using our experience from previous studies to develop content and design that will appeal to the target populations, is culturally sensitive, and successfully communicates the importance of study participation. All materials, including the questionnaires, will be developed in both English and Spanish.

We recognize the challenges involved in reaching members of ethnic and racial minority groups who may be difficult to interview. Survey participation on the part of Native Americans and Hawaiians is of particular concern because these groups tend to be more distrustful of Western medicine and less familiar with the clinical institutions and survey procedures. However, we expect that the experience of completing the baseline CSA will reduce their reluctance to complete a subsequent beneficiary survey.

To maximize response rates, we will give a $20 incentive (in the form of a check or gift card) to survey respondents. Several other techniques will also be employed to enhance response. First, we will develop a Frequently Asked Questions (FAQ) brochure for the mailings that addresses specific concerns of the target populations. Second, a toll-free telephone number will be set up to address questions from sample members. Third, we will permit proxy respondents to complete the questionnaire in cases in which the sample member is too ill or impaired to do so.

To ensure that we obtain current addresses and telephone numbers for sample members, we will request telephone numbers from a commercial telephone number look-up service and from the Social Security Administration. We will also use expert telephone tracers in RTI's in-house Tracing Operations (TOPS) unit for limited intensive tracing activities. We also plan to approach the demonstration sites for assistance locating participants when necessary.

In each study arm, we project that response rates will be higher in the intervention group (73%) than in the control group (65%). Based on these projections, we will contact 7,246 beneficiaries to obtain 5,000 completed interviews.

4.5   Analysis Plan

We will conduct a series of statistical analyses of the beneficiary survey data to assess the impact of the facilitation services on beneficiary satisfaction and health outcomes.

Descriptive and reliability analyses. We will begin our analyses by computing descriptive statistics (means, standard deviations, and frequency distributions) for all key survey outcome variables. Many survey outcomes, such as the SF-12 physical and mental health component scores, have established scoring procedures. We will check the internal consistency of multi-item measures like the CAHPS scales using Cronbach's alpha.

Response propensity analysis. Once data collection has been completed, we will conduct an analysis of response rates to determine whether survey participation was influenced by intervention status or other background characteristics. This response propensity analysis is based on a logistic regression model in which the binary outcome is coded 1 if the sampled beneficiary completed the baseline and 0 if the beneficiary did not complete a survey. The explanatory variables in the model will consist of factors that are available from the baseline CSA for all beneficiaries, such as demographic characteristics, chronic disease diagnoses, and intervention status, as well as chronic disease diagnoses from claims data. Factors contributing to differential participation will be incorporated in the sample weights for the study to adjust for non-response bias.

Analysis of intervention effects. In this randomized study, we will employ weighted analysis of covariance regression models to estimate the effect of the CPTD interventions on beneficiary satisfaction and health outcomes scores. The general model for the intervention analyses is:

$Y = a + b_1 Z + b_k X_k + e$
where:
Y = a follow-up outcome measure,
Z = an intervention status indicator (1=intervention, 0=control);
X_k = a vector of k covariates,
b_1, and the b_ks are regression coefficients to be estimated,
a = an intercept term, and
e = an error term.

In this model, coefficient b_1 estimates the overall effect of the intervention on the designated outcome. The covariates in the model will be drawn from characteristics known to influence reported health outcomes such as age, gender, educational attainment, and cancer stage. These covariates help to increase the precision of the estimated intervention effect. A key covariate will be the baseline CSA measure of each outcome. In addition, due to the rolling recruitment procedures that sites will be using, we will also control for the number of months elapsed between the completion of the baseline and follow-up surveys.

Because the intervention services provided in each site are expected to be unique, models in this general format will be estimated separately for each site and for each arm of the study. All intervention subjects will be included in the intervention group, regardless of their level of participation in this intent-to-treat analysis.

Figures 1 and 2 show a general analysis model for the screening and treatment arms that adds some claims-based outcomes. We expect that the facilitation intervention will have its greatest direct impact on access to services, adherence to guidelines, and satisfaction with care.

An intervention effect may also be transmitted indirectly to some outcomes through its impact on these variables.
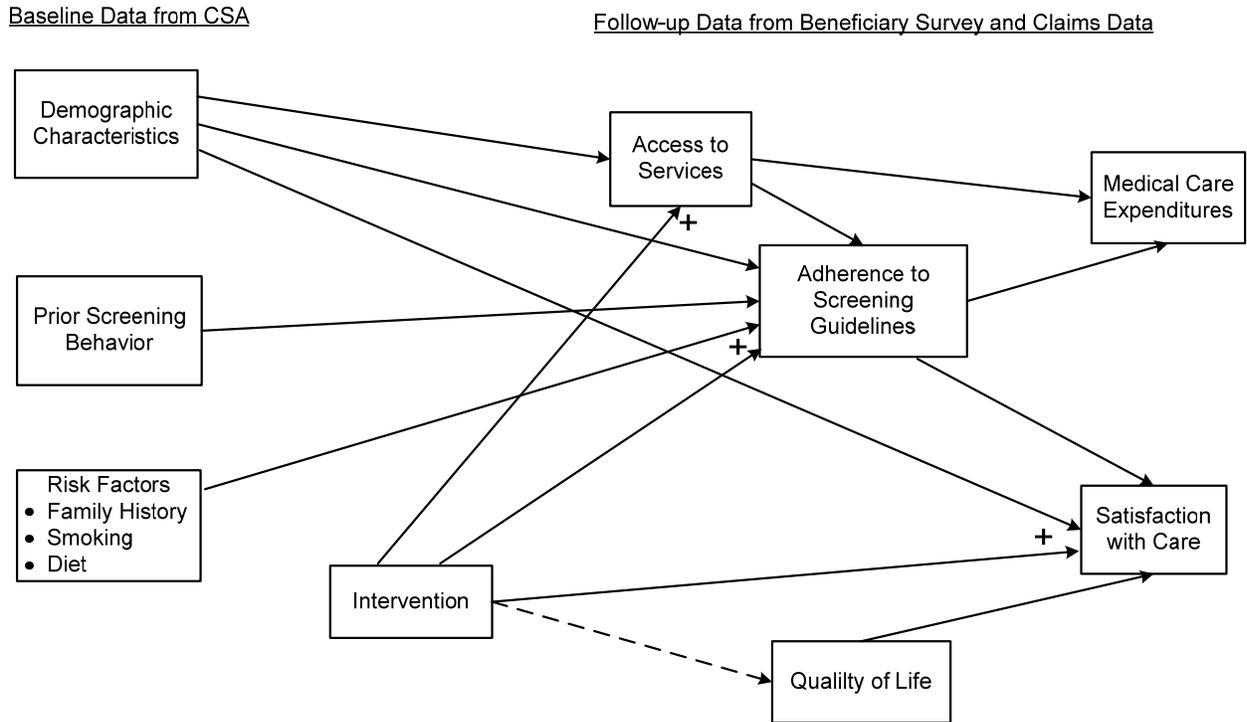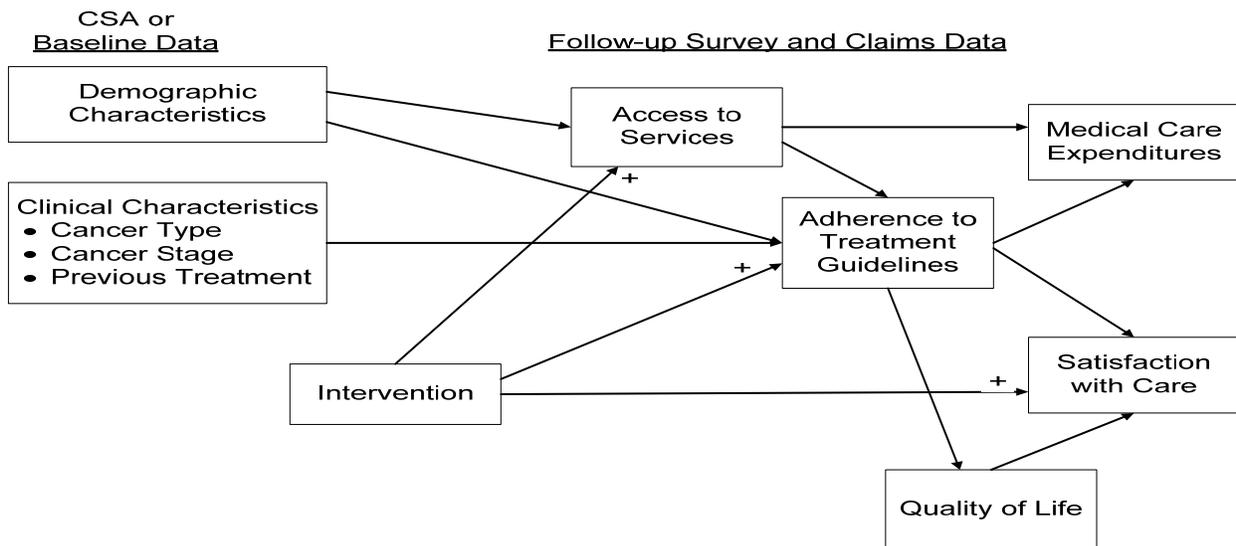
Figure 1
Analysis Model for Screening Arm

Baseline Data from CSA

Follow-up Data from Beneficiary Survey and Claims Data

Figure 2
Analysis Model for Treatment Arm

CSA or
Baseline Data

Follow-up Survey and Claims Data

4.6   Baseline CSA Analysis

Many of the items in the baseline CSA are standardized measures that have previously been administered to the general population in surveys like the cancer supplement to the National Health Interview survey. This provides an opportunity to compare the responses of the minority groups targeted by the CPTD demonstration to those reported by the general population. We will prepare a set of tables showing the mean values for selected CSA measures by site and for the reference population. The purpose of these analyses is to document disparities that are present **prior** to the implementation of the facilitation interventions and to identify variations between the individual sites in these measures.

Screening Arm. In this arm, we will compare site-specific data to nationally representative rates for older adults for the following measures:

- Self-reported screening rates for each of the five target cancers.

- Reasons for not having a particular cancer test.

- Cancer knowledge, beliefs, and perceived susceptibility (CSA items B1-B6).

- Physical and mental functioning (SF-12 component scores).

- Health utilities (EQ-5D).

Treatment Arm. The two primary cancer-specific measures in the CSA are the FACT-G functioning scales and satisfaction with care. Nationally representative data for cancer patients are not available. Instead, we will compare the rates reported by CSA respondents to selected large surveys for the target cancers that included these scales. The following measures will be used in the tables:

- Overall functioning (FACT-G total score).

- Satisfaction with care (EORTC QLQ-SAT32).

- Health utilities (EQ-5D).

SECTION 5
MEDICARE CLAIMS

5.1 Data Sources

5.1.1 Demonstration Participant Data

The Medicare claims analyses will draw on Part A MedPAR data and the Part B Standard Analytic Files (SAFs), as well as beneficiary information in the Medicare denominator file. We will extract claims for members of the intervention and control groups using Medicare HIC number. If they are available, we will also include Part D prescription drug event data for those demonstration participants who are enrolled in a prescription drug plan. We will use claims data to identify both service quantities and Medicare allowed charges.

Medicare administrative data will be supplemented by person-level data provided by the sites. We expect that these person-level data will be collected as part of the initial CSA, as well through the annual CSA for the intervention group and through the exit CSA for the control group. We will link the supplemental person-level data to the Medicare administrative data using HIC numbers. Because screening and treatment protocols often vary by individual characteristics, additional clinical data may be required to identify appropriate screening and treatment standards for each member of the study population. Furthermore, compliance and the effectiveness of the interventions may vary across subpopulations. For example, marital status and whether the beneficiary lives alone may affect compliance with screening and treatment recommendations.

We expect that the initial CSA will include the following information that will be incorporated as explanatory variables in the claims analyses or used to construct clinically appropriate outcome measures:

- Marital status

- Living situation

- Number of living children

- Language

- Income

- Education

- Usual source of care

- Screening history (screening arm only)

- Cancer risk factors (screening arm only)

- Date, type, and stage at diagnosis (treatment arm only).

If the data are available, we would also incorporate information on screening results and diagnosis following an abnormal screening result for members of the screening arm. These data could be collected for the intervention group through the annual CSA and retrospectively at the end of demonstration participation through the exit CSA for the control group.

5.1.2 Baseline Data

Baseline disparities in cancer screening will be examined using outpatient SAF data and NCH Part B data for the two calendar years immediately preceding the demonstration: 2004-2005. These claims data will be linked to denominator file in order to obtain beneficiary characteristics (including race/ethnicity) and to identify any periods of managed care enrollment.

Baseline disparities in cancer treatment will be based on the SEER-Medicare Linked Database maintained at the National Cancer Institute. The Surveillance, Epidemiology and End Results (SEER) programs collects data from selected population-based cancer registries around the country. These data include clinical and demographic information on all persons diagnosed with cancer in the geographic areas they cover, including date of death. We will be using the most recent, and greatly expanded, SEER data that represents 26 percent of the U.S. population diagnosed between 2000 and 2002. The linked database includes the Medicare enrollment database and all Medicare claims for these cancer patients from 2000 through 2003.

5.2   Baseline Analyses

The CPTD was designed to assess the impact of facilitation services on cancer screening and treatment outcomes for minority group members, and thereby reducing pre-existing disparities between minority and white Medicare beneficiaries. Each of the funded demonstration projects is limited to a single minority group. In order to strengthen the experimental design, participants receiving the intervention will be compared with randomly assigned participants in the control group (with members of both groups belonging to the same minority group). The experimental design does not include direct comparisons with whites. However, without knowing the magnitude of the disparity between racial/ethnic minorities and whites to start with, the evaluation will not be able to determine how much of the gap was closed.

To address this, we will conduct baseline analyses to identify differences between racial/ethnic minorities and whites in utilization of cancer screening and treatment services, as well as differences in Medicare costs for these services. In the following two subsections, we first describe our plan for analyzing baseline disparities for the population receiving screening services and then describe our plan for the treatment population.

5.2.1 Medicare Analysis of Screening Disparities

The analysis of baseline screening disparities will use Medicare claims data to examine utilization of screening services for breast, cervical, colorectal, and prostate cancer, comparing whites with the four racial/ethnic minorities targeted by the demonstrations (African-Americans, Hispanics, Asian/Pacific Islanders, and American Indian/Alaskan Natives). Comparisons between whites and each of minority population will be made on a national level. In addition, in the geographic areas where demonstrations are located, we will compare whites with the population that is the focus of that demonstration. The analyses will be based on Medicare Part B

claims for the two years prior to implementation of the demonstration (2004-2005). Claims will be linked with information on beneficiary characteristics from the denominator file.

The baseline analyses of screening disparities will address the following issues:

- Does adherence to cancer screening guidelines differ for whites and racial/ethnic minorities?

- If there are differences, does the magnitude vary by type of screening test?

We will develop indicators for receipt of screening services using procedure codes in Medicare Part B claims data.[1] In collaboration with our oncology consultant, we will also develop algorithms to measure compliance with practice guidelines for each screening test. Since screening guidelines sometimes vary across organizations, as appropriate we will measure compliance using several sets of guidelines and will test the sensitivity of our results to the guideline used. Because we will not have information on individual risk factors, the algorithms will use age- and gender-appropriate standards for a person who is not high-risk. For example, some guidelines indicate that women over age 70 who are not at high risk can discontinue mammograms and pap tests. Therefore, our algorithms for mammograms and pap tests might differ for female beneficiaries who are younger and older than age 70. To the extent that the distribution of risk factors differs between whites and racial/ethnic minorities, the results of these analyses will be biased. If racial/ethnic minorities are more likely than whites to be high risk, then these comparisons will provide a conservative estimate of disparities in compliance with screening guidelines. As discussed in Section 5.3, this limitation will be addressed in the demonstration evaluations by incorporating data on individual risk factors obtained from the CSA.

5.2.2 SEER-Medicare Analysis of Treatment Disparities

The proposed SEER-Medicare analysis will provide baseline comparisons of cancer treatment for racial/ethnic minorities with whites, including both primary treatment and appropriate adjuvant treatment. The analysis will be limited to the five cancer sites that are the focus of the treatment arm of the demonstration: breast, cervix, colorectal, lung, and prostate. The following groups will be studied: white, African-American, Hispanic, Asian/Pacific Islander, and American Indian/Alaska Native. The latter group will be included to the extent that there are sufficient patients with each of the study cancers. We will use the most recent three years of SEER data (patients diagnosed in 2000–2002), with Medicare claims through 2003.

Because the dataset is greatly expanded for these most recent years, this will help maximize the number of cases for each of the minority groups and for cancers with relatively low incidence rates (e.g., cervix). SEER sites for this time period include nine entire states (California, Connecticut, Hawaii, Iowa, Kentucky, Louisiana, New Jersey, New Mexico, and Utah), and four metropolitan and rural areas (Atlanta, Detroit, Seattle-Puget Sound, and rural Georgia). The SEER registries also include American Indian/Alaskan Natives from Alaska and

---

[1] As appropriate, following CMS practice, we will use procedure codes for both screening and diagnostic tests to measure screening rates.

Arizona; we exclude these two groups from our analyses, as there are no white cancer patients from those two states for comparison. The SEER sites in our study represent 26.2 percent of the U.S. population, including 23.4 percent of the white population, 22.7 of African-Americans, 40.2 percent of Hispanics, 53.3 percent of Asian-Americans, and 27.9 percent of American Indian/Alaskan Natives (http://seer.cancer.gov/registries/data.html).

Comparisons will be made across all SEER sites pooled, and for those geographic areas in which the SEER site overlaps with one of the demonstration project sites. Comparisons with whites will be conducted separately for each of the four minority groups. For example, treatment for African-American and white cancer patients would be compared for the U.S. as a whole, as well as for Detroit. We would not be able to make African-American-white comparisons for the second African-American demonstration site of Baltimore, since that city is not included in the SEER registries.

To the extent sample sizes permit, comparisons will be cancer site-specific as well. However, we expect that many analyses will need to pool across types of cancer. The number of Medicare minority group members with a given cancer may be small for several reasons. First, life expectancy for minority groups is lower than that for whites, particularly among Native Americans/Alaskan Natives. As a result, there may be a relatively small number of individuals aged 65 and older. Second, at least one of the cancer types (cervix) is relatively low incidence. Third, the restriction of many demonstration sites to a single city will reduce the absolute number of cancer patients available for study.

Research questions include:

- What is the difference between whites and racial/ethnic minorities in completion rates for primary treatment?

- What is the difference between whites and racial/ethnic minorities in completion rates for multi-modality treatment (surgery plus subsequent chemotherapy and/or radiation therapy)?

- What is the time interval from start to finish of treatment for white vs. racial/ethnic minorities?

- What is the difference between whites and racial/ethnic minorities in post-treatment follow-up (surveillance) rates?

Treatment standards will be based on clinical practice guidelines (e.g., National Comprehensive Cancer Network), as defined for each type of cancer and stage at diagnosis. Type of cancer and stage at diagnosis will be obtained from SEER data. Treatment will be operationalized using procedure codes on Medicare claims. We will work with our oncology consultant (Dr. Earle, who has experience with SEER-Medicare data) to develop and validate all of our algorithms. Outcome measures for cancer patients within each of the four racial/ethnic minority groups will be compared separately with those for white cancer patients, and tested for statistical significance. Logistic regression will be used to control for sociodemographic characteristics, such as age, gender, original reason for Medicare entitlement, and dual eligibility

status. as well as stage at diagnosis. For those analyses pooling cancer types, we will include dummy variables for type of cancer. The odds ratios associated with the respective variables indicating minority group status will be used to determine whether there is a significant treatment disparity and, if so, the magnitude of that disparity.

In addition, we will compare baseline treatment costs for white cancer patients and those in the four racial/ethnic minority groups. To the extent that minority cancer patients are less likely to complete standard treatment protocols, their total Medicare costs may be lower than those of white patients. If so, this will provide useful insight as to the change in Medicare expenditures we may observe under the demonstration.

5.3   Analysis Plan for Evaluation

Table 2 summarizes the evaluation questions that will be addressed using Medicare claims data. Research questions for the screening arm include:

- Did the screening intervention increase screening rates and adherence to cancer screening guidelines?

- Did the screening intervention increase the likelihood of reaching a definitive diagnosis for a beneficiary with an abnormal screening result? Did the screening intervention increase the likelihood of adherence to diagnostic guidelines? Did it reduce the time until a definitive diagnosis was reached? Alternatively, did the time until a definitive diagnosis was reached increase because people in the intervention arm are identified earlier in the disease process and more testing is required to reach a definitive diagnosis?

- Did the screening intervention increase the likelihood of early cancer diagnosis?

Research questions for the treatment arm include:

- Did the treatment intervention increase primary treatment completion rates?

- Did the treatment intervention increase completion rates for multi-modality treatment?

- Did the treatment intervention decrease the time to completion of treatment?

- Did the treatment intervention increase the likelihood of adherence to treatment guidelines?

- Did the treatment intervention increase the likelihood of receiving post-treatment surveillance?

Treatment arm analyses will adjust for stage at diagnosis.

For both the screening and treatment arms, we will also look at the following issues:

- Did the intervention increase use of Medicare services not related to cancer screening and treatment? If so, what types of services?

- Did the intervention increase Medicare spending? If so, what types of services contributed to this increased spending?

We will use procedures codes in Medicare claims data, supplemented with patient-level clinical data obtained from the demonstration sites through the initial, annual, and exit CSAs, to construct variables that capture utilization of screening and diagnostic follow-up services for members of the screening arm and utilization of treatment services for members of the treatment arm.[2] Screening and treatment standards will be based on the specific guidelines adopted by each site, as well as cancer practice guidelines published by nationally recognized organizations[3]. While the guidelines adopted by the sites are certainly important metrics against which there performance should be evaluated, it will also be valuable to evaluate them in comparison with standard, nationally recognized guidelines. As described previously, we will work with our oncology consultant to create individual algorithms for each screening test and each type of cancer based on their specific guidelines. While screening guidelines for breast and cervical cancer are fairly standardized, there is greater controversy around prostate and colorectal cancer screening. Furthermore, in some cases, Medicare does not cover the frequency of screening that some physicians recommend.[4] We will construct measures of screening compliance using several guidelines and will test the sensitivity of our results to the guideline used. We will make a final decision about which guidelines to use in consultation with the CMS Project Officer.

Analyses of Medicare claims data will be undertaken separately for the screening and treatment arms. Our analyses will focus primarily on differences between members of the intervention and control groups at a site. If possible, in the treatment arm we will conduct separate analyses by type of cancer within a site; however, we expect that in most cases the sample sizes for individual cancers will not be adequate to support these analyses. Therefore, we will construct general outcome measures that can be applied across all types of cancer. *Table 4* shows the dependent variables that will be constructed to evaluate each arm of the demonstration.

---

[2] As in our baseline analyses, we will follow CMS practice and will use procedure codes for both screening and diagnostic tests to measure screening rates.

[3] Guidelines for cancer screening, including a comparison of recommendations from different national organizations, are available from the National Guideline Clearinghouse (www.guideline.gov). Treatment algorithms are available from the National Comprehensive Cancer Network (www.nccn.org).

[4] For example, Medicare covers one colonoscopy every 2 years for high risk patients. However, some research suggests and physicians believe the follow-up should be annual, depending on the high risk factor.

We will identify appropriate screening tests for each demonstration participant in the screening arm based on age, gender, individual screening history, and family risk factors. Guidelines for the frequency of some tests differ by risk status. For example, a colonoscopy is only required every 10 years unless the individual is in a high-risk group, in which case it is required every 24 months.

If supplemental data on screening results and diagnosis following an abnormal screening result for members of the screening arm are available, we can more accurately target variables related to diagnostic testing to individuals with abnormal test results. If these data are not available, we will not be able to construct some of the more specific dependent variables for the screening arm shown in *Table 4*. Instead, we will construct modified versions of the dependent variables that take advantage of random assignment of individuals to the intervention and control groups, which should produce groups with similar characteristics on average. Because of randomization, any resulting errors in constructing these variables should affect both groups in the same way. For example, we will not be able to measure time from an abnormal screening result to initiation of diagnostic testing if we do not know which screening tests have an abnormal result. However, because we can assume that the proportion of abnormal results will be similar in the intervention and control groups, we would expect the rate of follow-up diagnostic testing to be the same in both groups in the absence of the intervention. Therefore, any observed differences in the rate of diagnostic testing can be attributed to the effects of the demonstration. Without additional clinical data, we cannot measure demonstration impacts on time from screening to a definitive diagnosis or stage at diagnosis for those with abnormal test results.

Table 4
Dependent Variables for Medicare Claims Analyses

Screening Arm
• An indicator for whether the individual received each of the appropriate screening tests
• For each screening test, an indicator for whether the frequency of testing complied with guidelines
• An indicator for whether an individual with an abnormal screening result received appropriate diagnostic follow-up tests
• Time from abnormal screening result to initiation of diagnostic testing*
• Time between diagnostic tests
• An indicator for whether an individual with an abnormal screening result reached a definitive diagnosis*
• Number of weeks from screening to a definitive diagnosis*
• Stage at diagnosis (for those with a definitive diagnosis)*
Treatment Arm
• An indicator for whether an individual receives surgery or other appropriate primary therapy
• An indicator for whether an individual receives chemotherapy and/or radiation therapy following surgery
• Time from date of primary treatment through completion of adjuvant therapy
• An indicator for whether an individual has post treatment follow-up visits to an oncologist or primary care physician

*Requires supplemental data from demonstration sites.

We will measure compliance with screening guidelines in several ways. Because it may be easier to achieve compliance with some screening tests than others (e.g., those that are noninvasive or that require a simple blood test), we will use several different measures to gauge compliance with screening guidelines. These include receipt of each individual screening test, the percent of tests received, and receipt of all appropriate tests.

Assuming the required data are available from the CSA, we will adjust outcome measures for the treatment arm to reflect an individual's stage at cancer diagnosis, treatments received prior to the initiation of the demonstration, and prognosis. If these supplemental data are not available, we will again assume that the distribution of cancer stage, prognosis, and prior treatments is similar across the intervention and control groups.

In addition to data on utilization of cancer screening and treatment services, for individuals enrolled in both arms we will construct variables for the utilization and cost of all Medicare services. Although we hypothesize that the interventions will decrease expenditures in the long run, they may increase over the short-run period covered by the evaluation. For both groups, in addition to total Medicare expenditures, utilization and expenditures will be disaggregated by type of service (inpatient, hospital outpatient department, physician/supplier, home health, prescription drugs, etc.). For the cancer treatment arm, utilization and costs will be further broken out by cancer-related and non-cancer-related services. For the screening arm, we will separately identify utilization and costs for cancer screening tests and other screening tests in order to identify whether the intervention increases awareness of the prevention and the use of screening services generally.

We will also compare outcomes for members of the intervention group, with a "pseudo" control group. Because of the nature of these demonstrations, there may be spillover effects from the interventions that lead to contamination of control group members. To estimate the extent of contamination, we propose comparing the control group with a "pseudo" control group that is identified using Medicare administrative data. The pseudo control group will be selected to be similar to each demonstration population based on characteristics available through the EDB (age, gender, race and ethnicity, Part D enrollment, dual eligibility status, original reason for Medicare entitlement, and geographic location). We will then pull claims for individuals selected for the pseudo control group. If they are comparable, utilization of screening and treatment services by this pseudo control group should be similar to that of the control group. Differences between the groups would be assumed to reflect the effects of contamination. To the extent we find evidence of contamination, comparisons between the intervention group and the pseudo control group provide alternate estimates of intervention effects.

Limitations of the race and ethnicity data in the EDB are well recognized. As a result, depending on the population targeted by the demonstration site, it may be difficult to identify a pseudo comparison group with the same racial and ethnic characteristics using the variables currently available in Medicare EDB. CMS has been focusing considerable effort on improving the quality of race and ethnicity data. RTI has developed an algorithm for CMS that improves identification of certain racial and ethnic minorities, particularly Hispanics. Using this algorithm, RTI has created a file, updated through October 2005, that identifies the racial and ethnic characteristics of Medicare beneficiaries. We will link this file to the EDB using beneficiary HICNO. We will also explore applying the algorithm to beneficiaries enrolled after October

2005, who are not included in the most recent update file. In addition, as an alternative check for contamination, we will compare utilization of screening services by control group members during the demonstration with a baseline period prior to initiation of the demonstration. (This strategy is not feasible for control group members in the treatment arm since their treatment needs presumably change over time.) If they continue to receive usual care and are not contaminated by the interventions, utilization by control group members should not change substantially over relatively short periods of time.

Our basic analytic model takes advantage of the random assignment of demonstration participants to the intervention and control groups in each of the arms. Because of this random assignment, the characteristics of the intervention and control groups should be comparable, and differences between the groups should reflect the effects of the interventions. We will conduct X^2 tests for discrete dependent variables. For duration variables (e.g., time from an abnormal test result to definitive diagnosis), we will calculate the Kaplan-Meier empirical hazard function and the corresponding nonparametric survival function. The hazard and survival functions take into account right-hand censoring of the data due to the end of the demonstration period. If sample sizes permit, we will conduct separate comparisons for subsets of the population defined based on characteristics that might be expected to affect outcomes (e.g., gender, age group, original entitlement to Medicare based on disability, income, language, marital status, living situation, having a usual source of care). While a few of these characteristics are available in the Medicare EDB, others would be based on supplemental data collected through the CSA.

In addition to descriptive statistics, we will also conduct multivariate analyses. The procedure used to estimate the multivariate models will depend on the nature of the dependent variable (logistic regression or Cox proportional hazard). The basic multivariate model is:

$Y = a + b_1 Z + b_k X_k + e$

where:

Y = a follow-up outcome measure,

a = an intercept term,

b_1, and the b_ks regression coefficients to be estimated,

Z = an intervention status indicator (1=intervention, 0=control)

X_k = a vector of k covariates,

e = an error term.

In this model, coefficient b_1 estimates the overall effect of the intervention on the designated outcome. The covariates in the model will be drawn from sociodemographic and other characteristics known to influence reported health outcomes such as age, gender, educational attainment, and cancer stage at diagnosis. These covariates help to increase the precision of the estimated intervention effect. In addition, they allow us to control for any imbalances in the characteristics of the treatment and control groups that might arise despite

randomization. Imbalances are particularly likely to occur if the population enrolled in the demonstration is relatively small. If sample sizes permit, this model can be extended by interacting Z and X_k to identify differing impacts of the demonstration on various beneficiary subpopulations.

In addition, we will consider models that combine data across sites. However, we believe that pooled analyses should be undertaken very cautiously because of significant differences in the target population across sites and potential differences in the interventions used. Any models that combine data from multiple sites would include site-specific variables to measure intervention effects.

SECTION 6
COST-EFFECTIVENESS, COST-UTILITY, AND BUDGET IMPACT ANALYSIS

6.1   Purpose

Cost-effectiveness is an essential component of the evaluation as costs of the facilitation services need to be assessed in relation to the benefits provided. Analysis focused on utilization and cost alone does not recognize the potential benefits of increasing screening participation, improving quality of life, and reducing mortality (Anderson et al., 2002; Gold et al., 2003; Haddix et al., 2003). Providing screening and treatment facilitation will generally result in additional costs but it is anticipated that this higher cost will be offset by improvements in the effectiveness of the care delivered and reductions in cost along the cancer care continuum. Therefore, it is essential to perform a systematic assessment of both the cost and effectiveness of the services provide in the demonstration. We will perform cost-effectiveness and cost-utility analysis of the facilitation services offered for screening and treatment.

In addition, we will also perform a detailed assessment of the cost incurred by the demonstration sites to provide the facilitation services (depending on data provided by the sites). This information will provide CMS with the cost of specific activities performed by the sites for the screening and treatment intervention groups relative to their control groups. Increasing screening rates and improving compliance may reduce health care costs in the long run, but these savings will not be fully realized during the four years of the demonstration. Long-term follow-up of the intervention and control groups beyond the demonstration period is necessary to fully capture all the benefits of the screening facilitation services provided. We will attempt to estimate the long-term impact of the facilitation services, by using modeling techniques to extrapolate the costs and benefits beyond the time period of the demonstration.

In summary, the demonstration costs and effectiveness will be analyzed to address the following research questions:
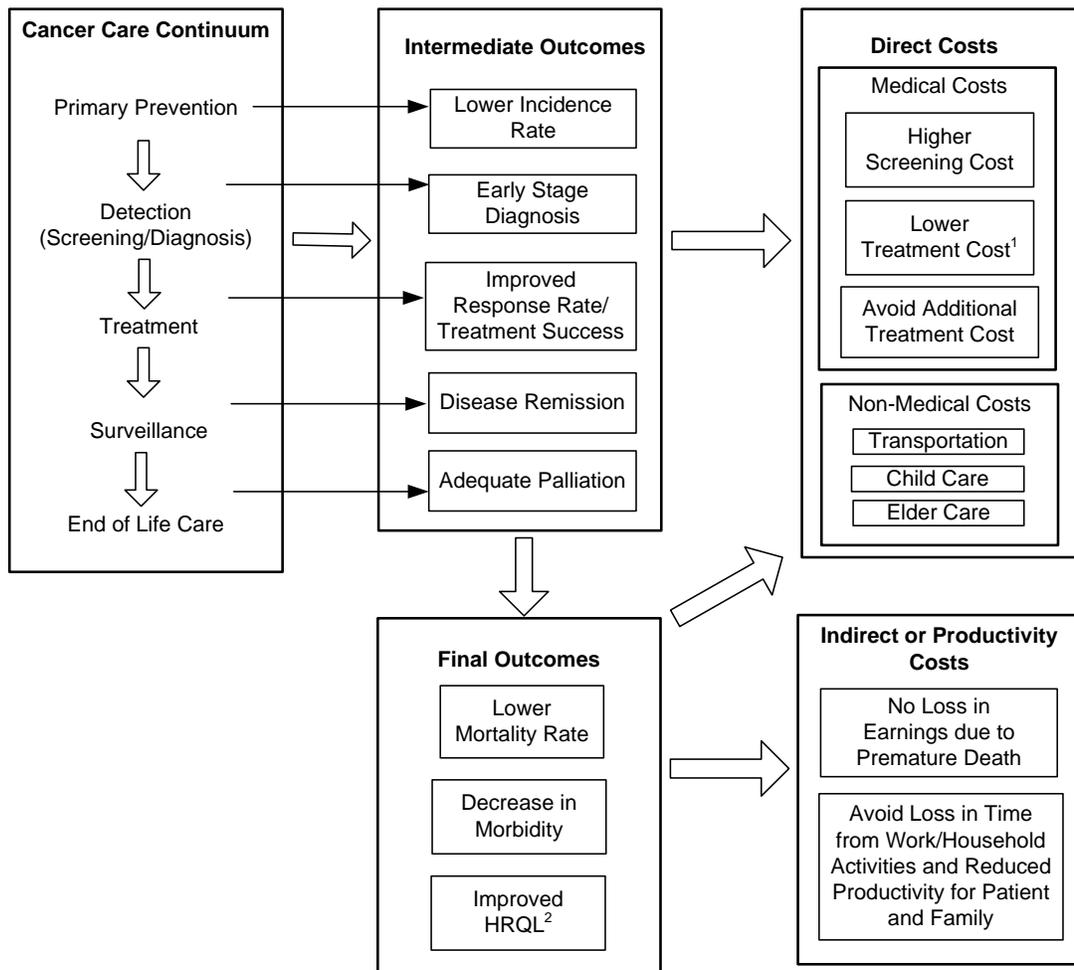
1. What is the cost of the facilitation services offered by the demonstration sites?

2. What is the average facilitation cost per person in the screening and treatment intervention groups?

3. What is the incremental cost-effectiveness and cost-utility of the facilitation services, that is, the cost-effectiveness of the screening and treatment intervention groups relative to their control groups?

4. What is the long-term budget impact and cost-effectiveness/cost-utility of the screening facilitation services when results are extrapolated beyond the period of the demonstration?

In this sections that follow, we present an overview of the approach to assessing the economic impact of the demonstration, describe the cost estimates and effectiveness measures that will be derived and provide details on the specific types of analysis that will be performed to answer the research questions.

6.2   Approach and Data Sources

Figure 3 provides a framework for assessing the outcomes and cost of cancer-related interventions. Cancer is a chronic disease and therefore health care services are provided along a continuum of care: beginning with primary prevention, then screening and diagnosis, followed by treatment and surveillance, and finally end of life care. Interventions that affect any point in the continuum will have impacts further along the continuum of care, both in terms of outcomes and cost. The focus of this evaluation will be on assessing the impact of screening, diagnosis and treatment. We will not have the information required to assess the impact of primary prevention.

Figure 3
Framework for Assessing Economic Costs of Cancer

[1] Stage IV cancers may have lower cost than Stage II and III cancer
[2] HRQL – Health Related Quality of Life

40

6.2.1 Intermediate and Final Outcomes

Offering facilitation services for screening and treatment will result in improvements in intermediate outcomes, including earlier stage of diagnosis and higher rate of treatment success. By definition, the intermediate outcomes lead to final outcomes. Final health outcomes are the ultimate measure of the impact of any intervention and an increase in compliance with screening and treatment recommendations should result in the following benefits:

- Reduction in mortality;

- Decrease in morbidity

    – Less burden on family members and community

    – Improved ability to work; and,

- Increase in health-related quality of life (HRQL).

During the demonstration period, only the intermediate outcomes will be observable for the intervention group and these will be quantified through analysis of the Medicare claims data and beneficiary survey. For the control group, even these intermediate outcomes will not be captured since cancer in this group may remain undetected for years. For example, if no colorectal cancer screening is performed, then polyps that may be present will not be removed, and the individual in the control group will present with late stage colorectal cancer related symptoms years after the end of the current demonstration. Therefore, to adequately compare the outcomes in the intervention group to the control group, extrapolation of events beyond the timeframe of the demonstration period is required. In addition, such extrapolation will allow for the assessment of the impacts of facilitation services on final health outcomes, which can then be compared with the cost-effectiveness of other health care interventions.

6.2.2 Perspective for Economic Assessment

There are several perspectives or viewpoints that can be used to perform economic assessments. For instance, the evaluation can be performed from the Medicare/CMS or payer perspective, the program perspective, or the societal perspective. The societal perspective is the most comprehensive and includes all the costs identified in *Figure 3*, that is, both the direct (e.g., health care services, child care) and indirect costs (e.g., work loss). The societal perspective will not be used in this evaluation because of the methodological challenges and the high cost of obtaining all the costs and benefits involved.

Since the findings from the demonstration will guide decisions regarding the future implementation of facilitation services for Medicare beneficiaries, analysis from a CMS or payer perspective will be performed. The payer perspective tends to be narrow, however, and may not reflect the true cost of replicating the facilitation activities. For instance, the CMS payments to the sites for providing the facilitating services may be lower or higher that the true cost incurred by the sites to provide these services. Therefore, we will also conduct the evaluation using the program perspective (i.e., including the actual costs of providing demonstration services). The program perspective will provide CMS with an understanding of the resources required to

implement similar programs in the future. In addition, this information will allow CMS to better estimate the cost of supporting more or less intensive facilitation services.

6.2.3 Activity-Based Program Costs

Collecting activity-based costs, which is the approach in which activities are identified and all related costs of performing them are systematically calculated, will allow CMS to assess the true cost incurred by the demonstration sites, identify factors that impact cost, and perform cost-effectiveness analysis from the program perspective. In addition, obtaining detailed cost by activity will help CMS to understand differences in the types of facilitation services offered across demonstration sites. Any detailed comparisons of the cost of specific activities will only be performed between sites that provide similar facilitation services.

We will use a previously tested questionnaire (based on site participation), the Cost Assessment Tool (CAT) to collect detailed activity based costs from the demonstration sites. The CAT is based on standard well-established methods for cost data collection (French et al., 2004; French et al. 1997; Salome et al., 2003; Zarkin et al., 2001) and has been administered successfully to collect costs from cancer programs. The CAT is an Excel based data collection tool in which information will be gathered electronically from the demonstration sites to eliminate data entry errors The sites can also check and correct their data inputs prior to submission. We will tailor the information collected in the CAT to obtain cost data on the facilitation services provided by the demonstration sites. We will provide "drop-down" menus as appropriate to reduce data input burden and develop a detailed user's guide (which will contain specific examples) to assist the sites in providing the cost information requested. OMB clearance will not be required to administer the CAT, as only program level costs will be collected from the 6 sites (no individual patient level data are required).

RTI will train the sites and provide technical assistance for completing the CAT. We will introduce the cost collection activities during the meeting with the sites in May and will identify a key contact person at each of the 6 sites for the cost data collection. Based on prior experience, we anticipate that three conference calls (between RTI and the sites) will be required. During the first call, RTI will train the sites on completing the CAT by reviewing the information to be provided and offering guidance on completing the requested data elements. Examples will be provided and the user's guide will be introduced. The second conference call will be scheduled after the sites begin to complete the CAT to answer questions that may arise. The third and final conference call will be conducted individually with each site after the CAT has been completed and provided to RTI so that any issues that arise during RTI analysis of the data can be discussed. (Depending on the timing of the initial site visits, training also could be conducted during those visits.) We anticipate that the sites will require 3 hours for training (review user's guide and participate in RTI training) and then on average10 hours for completing the CAT. We are proposing that the sites complete the CAT at 4 different intervals during the demonstration (as explained below) and therefore the total burden for the sites will be 43 hours. The greatest burden will be at the start of the demonstration when training and the first round of data collection will occur. Subsequent rounds of data collection will require less effort as the sites will have prior experience in providing the requested data and as RTI will be using automated processes to analyze the data.

We will collect costs associated with the start-up activities, as well as ongoing program activities, to assess the potential cost of implementing similar programs in the future. The information collected will provide details on both the fixed and variable components of the facilitation services. *Table 5* presents the types of data elements to be collected both in the start-up and implementation phases. A component of the CAT will be designed specifically to collect start-up costs (Start-up CAT). The start-up time period will be defined as the time period from the initiation of the demonstration to when the first beneficiary is enrolled in the program. The Annual CAT will collect information on implementation costs for each of the first three years of the demonstration. There can be anomalies in any given year that can impact the program costs and therefore having three years of data will provide a better estimate of the recurring costs. In both the Start-up and Annual CAT, we will collect information on in-kind contributions (donated labor and other resources) since these contributions can be a very significant proportion of the resources expended by cancer programs (Subramanian et al., 2005). The timeline for administering the Start-up and Annual CAT is presented in *Table 6*. The Start-up CAT will be completed at the end of the start-up period and the Annual CAT will be completed about 3 months after the end of each demonstration year (first three years) to allow time for all expenditures for the relevant period to accrue.

Table 5
Summary of Program Level Costs

Start-up Cost (Start-up CAT)	Implementation Cost (Annual CAT)
- Hiring staff - Buying equipment - Training - In-kind contribution	- Personnel - Consultants - Materials and supplies - Travel cost - Administrative Cost - In-kind contribution

CAT – Cost Assessment Tool

Table 6
Timeline for Program Cost Data Collection

Data Collection Tool	Year 1		Year 2		Year 3		Year 4	
Start-up CAT	X							
Annual CAT*			X		X		X	

* Cost data for the preceding year of the demonstration will be collected.

The CAT is specifically designed to limit data collection burden and therefore collects information on budget categories that are familiar to the respondents. The cost information collected will be allocated to specific program activities based on proportion of time or resources expended on the activities as reported by the respondents. For example, under personnel

expenditure, all program staff will be asked to provide the proportion of time spent on each of the program activities. These activities include:

- Program management

- Outreach/Recruitment

- Patient support/Navigation

    – Screening

    – Treatment

- Data collection and tracking

- Program evaluation

Whenever possible, we will attempt to identify the specific costs associated with screening and treatment facilitation. The approach to identify these costs will be finalized after discussion with the sites. If collecting this level of detail proves to be burdensome, we will use an alternative algorithm to allocate costs. For instance, we could use the proportion of individuals for whom screening versus treatment facilitation services are provided. In addition, we will also distinguish between the costs associated with services provided to the intervention versus the control group. For example, the control group enrolled by the *Molokai General Hospital* will receive a Nutrition Education Program and the cost associated with these services will be estimated to allow for an appropriate comparison. The approach for collecting costs associated with the services provided to the control group will be site specific and will depend on the type of service offered. The cost of the services provided to the intervention and controls groups will be assessed in relation to their respective effectiveness (discussed below).

The data submitted using the Start-up and the Annual CAT will undergo thorough data validation to assess the quality of the data available to perform the planned analysis. All data collected will be assessed for missing information (% of fields with missing data), and incorrect data (e.g., % data elements with formats that are not recognized; % with inappropriate range of values). We will also review whether the subcategories sum up to the expected total costs. Discrepancies between the total amount of funds expended annually and the total itemized costs will be identified and clarified with the sites. In-kind contributions will also be reviewed to ensure that only those contributions that represent true opportunity cost are included. Opportunity cost is defined as the "advantage forgone as the result of the acceptance of an alternative." For example, a person who volunteers his or her time will not be able to devote the time spent on the demonstration to other activities. The time spent should therefore be valued at the market rate and included as a cost to the program. The findings from the data validation will be reviewed to identify if any statistical or other corrections are required to generate unbiased cost estimates.

6.2.4 Type of Cost-Effectiveness Analysis

There are three types of approaches to simultaneously consider the cost and effectiveness of an intervention: cost-effectiveness analysis (CEA), cost-benefit analysis (CBA), and cost-utility analysis (CUA). In each of these three approaches, a ratio of the cost divided by the effectiveness units is generated and results are presented as a cost per unit of effectiveness (see *Table 7*). As appropriate, the effectiveness measures are reported in their natural units, in dollar terms, or as quality adjusted life years (QALYs). Cost-benefit analysis is rarely performed for cancer evaluations and is not recommended due to the challenges of quantifying both costs and benefits in monetary terms. Due to the chronic nature of the disease process and the anticipated impact on HRQL of the patients, cost-utility analysis is theoretically the most appropriate method for assessing cancer interventions (Anderson et al. 2004; Fishman et al, 2000). Cost-effectiveness analysis though is the most commonly reported ratio because of the difficulties in deriving unbiased measures of quality adjusted life years. For this evaluation, we will perform both cost-effectiveness and cost-utility analysis.

Table 7
Comparison of CEA, CBA and CUA

Type	Description	Cost Measure	Effectiveness Measure	Ratio
CEA	Consequences or effects of the intervention are expressed in natural units, such cases detected, cases successfully treated, or some other improvement.	Dollar	Natural units	Cost per unit life year gained
CBA	Both costs and benefits are expressed in monetary terms of net savings or a benefit-cost ratio. A benefit-cost ratio greater than 1 indicates that the intervention saves money.	Dollar	Dollar	Cost per $1 of benefit
CUA	Consequences are expressed as the utility or quality of the health outcome. CUA results are generally expressed as cost per QALY gained.	Dollar	QALY	Cost per QALY

6.2.5 Generating QALYs

The outcomes from screening and treatment have two basic components – the quantity and the quality of life. A QALY embraces both of these components and when measured over the lifetime it is the arithmetic product of life expectancy and a measure of the quality of the remaining life years. QALYS can also be generated for a given health state by multiplying the amount of time spent in that state (in years) by the quality of life during the health state. QALYs can be generated using the following formula:

$$QALY = \Sigma u_i * HS_i$$

where u_i = the utility weight and HS_i = the duration of time in a given health state i

A year of perfect health is worth 1 and a year of less than perfect health is worth less than 1. Death is considered to be equivalent to 0, however, some health states may be considered worse than death and have negative scores. The advantage of QALYs is that the derived unit can be used to compare interventions with widely different effects on health, particularly those that affect life expectancy and HRQL in different ways. In addition, when combined with the cost of providing the interventions, cost–utility ratios can be generated which can again be used to compare different types of programs.

The QALYs for the demonstration participants will be calculated based on the utility values generated from the responses to the EQ-5D (See Section 4.0 for details). The values for the intervention and control groups will be based on the changes from baseline survey to the follow-up survey. Since we will only have two points of data collection, we will calculate QALYs to study the impact of the demonstration as the difference in the utility values between the baseline and follow-up response multiplied by the time between these two responses. The QALYs for the intervention and control groups of the screening and treatment facilitation services can be derived as follows:

$$QALYs = (U_F - U_B)*(T_F - T_B)$$

where U is utility value, T is time, B is the baseline survey and F is the follow-up survey. The incremental change in QALYs for the intervention versus control group will therefore be a difference-in-difference assessment. Despite random assignment, there may be differences in baseline values between the intervention and control groups, and our proposed approach will help adjust for this. In addition, having baseline values will provide the opportunity to assess the potential magnitude of change in utilities due to the facilitation services in both the screening and treatment arms.

6.3   Analysis plan

6.3.1 Cost of the Facilitating Services

We will perform a thorough assessment of the cost related to offering the facilitation services using the detailed program level cost data collected from the sites using the CAT. We will report the cost of patient support and navigation services for screening and treatment facilitation separately. In addition, we will allocate other costs, including program management, outreach activities, data collection and tracking, and program evaluation, using information collected from the sites for screening and treatment facilitation. Any cost associated with offering services to the control group will also be reported.

Summary statistics will be generated for these costs and compared across the demonstration sites. We will show the possible range of values and generate univariate statistics (e.g., mean, standard deviation, median, interquartile range). Total cost and cost for the individual components, as applicable, will be compared and we will develop histograms to facilitate the comparison. We will also generate the average per person cost of facilitation for screening and treatment services for each site. In order to assess potential economies of scale (projected cost for future programs with differing screening and treatment volumes), costs that

are fixed versus variable will be identified for each site. Fixed costs when amortized across large number of screens could significantly decrease cost.

In addition, we will also report the magnitude of in-kind services received by the sites and the types of activities made possible by these contributions. The cost for the facilitation services obtained from the demonstration sites will be compared to the payment provided by CMS. We will also compare the cost of facilitating services reported by the demonstration sites to those reported by other programs (Saywell et al. 2004; Crane et al. 2000; Lynch et al. 2004).

6.3.2 Incremental Cost-Effectiveness and Cost Utility Assessment of Facilitating Services

The incremental cost-effectiveness of the facilitation programs will be assessed by comparing the difference in effectiveness and cost between the intervention and control groups. The incremental cost-effectiveness will be calculated for both the screening and treatment facilitation as follows:

$$\frac{\text{Cost}_{\text{Intervention Group}} - \text{Cost}_{\text{Control Group}}}{\text{Effectiveness}_{\text{Intervention Group}} - \text{Effectiveness}_{\text{Control Group}}}$$

The effectiveness measures for screening facilitation will include the total number of persons screened or served, the total number of screening tests performed, and the proportion of abnormal tests receiving both timely and appropriate follow-up. The effectiveness measure that will be used to assess treatment facilitation is the proportion of beneficiaries receiving appropriate and timely treatment services. QALYs will be assessed for both screening and treatment facilitation.

Table 8 provides an overview of the cost-effectiveness measures, the patients included in each measure, and the data source. The cost-effectiveness ratios will be derived using the facilitation cost from the CMS and program perspective. We will generate incremental cost-effectiveness ratios for each of the first three years of the demonstration separately and pooled together. We will perform nonparametric bootstrapping to evaluate the uncertainty of the results from the cost-effectiveness calculations to generate 95% confidence intervals. We will compare the results among the demonstration sites when appropriate (comparisons will be made among programs with similar facilitation services). In addition, the incremental cost-effectiveness of the facilitation services for screening and treatment will be compared to other published cost-effectiveness of health promotion programs (Fishman et al., 2000; Lynch et al., 2004; Saywell et al., 2004). The cost per QALY will be compared to league tables (Chapman et al. 2000; Mauskopf et al, 2003) reported in the literature. These tables compare the relative cost-utility of different interventions and provide a listing of their incremental cost per QALY. The ratios derived from the proposed analysis will be compared to those provided in league tables to understand the relative cost-effectiveness of the facilitation services. Specific attempts will be made to compare with other cancer screening and treatment facilitation services, but due to differences in methods used and type of population studied, the results may not be directly

Table 8
Cost-Effectiveness Ratios and Data Sources

Measure	Patients	Data
Cost per person screened or served Cost per screening test performed	All receiving screening	Program Data (CAT) Medicare Claims
Cost per person receiving appropriate and timely diagnosis	Those with abnormal test results	Program Data (CAT) Medicare Claims
Cost per person receiving appropriate and timely treatment	All receiving treatment	Program Data (CAT) Medicare Claims
Cost per QALY (screening & treatment)	Sample responding to survey	Program Data (CAT) Medicare Claims Survey Results

QALY – Quality Adjusted Life Year

CAT – Cost Assessment Tool

comparable (Briggs et al., 2000). Overall, the facilitation services will be considered cost-effective if the values are lower than the generally accepted threshold of $50,000 per QALY.

6.3.3 Long-Term Budget Impact and Cost-Effectiveness Assessment

We will use both program costs obtained using the CAT and the patient level costs derived from Medicare claims (details provided in Section 5.0) to estimate the cost of screening, facilitation, and clinical services from both the CMS and program perspective. The cost associated with the clinical services (screening, diagnosis, treatment and surveillance) will remain the same for either perspective but the cost associated with the facilitation services will differ. For the CMS perspective, we will include the capitation payment provided to the demonstration sites (allocated to screening versus treatment intervention based on numbers enrolled) and for the program perspective, we will include costs based on the activities performed for screening facilitation.

We will develop a simple excel based model to project the costs beyond the timeframe of the demonstration to assess long-term budget impact. The costs incurred during the first three years of the demonstration for the intervention and control groups will be extrapolated over the lifetime of the two cohorts and along the entire continuum of care indicated in *Figure 3*. This model will include epidemiologic parameters such as average life span of Medicare beneficiaries and incidence of cancer. Since the groups are randomly assigned to either the control or diagnosis group we will assume that the incidence of cancer will remain the same for both groups. The clinical costs of screening and diagnostic tests, treatment, surveillance and end-of life care will be obtained from analysis of Medicare claims and also from the literature (Brown et al., 2002; Chang et al, 2004). We will also use the model to assess the impact of the facilitation services on final health outcomes and to perform cost-effectiveness assessment. Information on health states and QALYs will be derived from the literature (Fryback et al., 1997; Ness et al., 1998, 1999)

Based on the information available during the demonstration, we will also attempt to model the budget impact for each type of cancer separately. The cost associated with screening, diagnosis and treatment differ by the type of cancer and these differences can only be taken into account if modeled separately. Also, for colorectal cancer, the costs of the recommended tests (e.g., fecal occult blood test, sigmoidoscopy, and colonoscopy) differ and thus overall costs will be impacted by the selection of the initial screening tests. We will therefore assess the budget impact based on several different test selection combinations. For each type of cancer screening performed, we will model two different scenarios: (1) the demonstration as a one-time effort with no facilitation services offered after the completion of the demonstration, and (2) the facilitation services continued beyond the demonstration. Under the second scenario the cost associated with screening facilitation will be incurred throughout the life span of the beneficiary. We will perform sensitivity analysis to understand the variation in the results obtained due to model assumptions.

SECTION 7
REPORTS

7.1 Reports to Congress

We will work with the Project Officer and other CMS staff as appropriate to develop the outline and format of each of the two Reports to Congress. All analyses completed by the due dates for the draft reports will be included. Timing of these reports is tied to the start of enrollment at the demonstration sites. The estimated due dates for the draft and final versions of the first Report to Congress are October 2007 and December 2007, respectively. The second Report to Congress will be submitted in draft on October 2009, and the final on December 2009. We describe the expected content of both Reports to Congress below.

2007 Report to Congress. This report will provide a comprehensive description of each of the six demonstration sites, including:

- Target population, recruitment methods, and catchment area;

- Intervention models for both screening and treatment arms;

- Payment methodologies;

- Enrollment to date in each group by age, gender, and other demographic characteristics, plus type of cancer (for the treatment arm);

- Baseline cancer screening history of participants; and

- Baseline health status and quality of life for participants.

Data sources will include the sites' original applications and subsequent revisions, progress reports, the first round of site visits, and analysis of all CSAs completed and submitted to CMS by late summer 2007.

Ideally, some preliminary analyses of Medicare claims for participants would be included in this report, but we suspect this will not be possible for several reasons. First, assuming a July 1, 2006 start date, there will not be sufficient time to observe utilization and costs unless a large number of participants are recruited in the first few months. A full year of observation will be needed to assess differences in screening rates. Second, enrollment of demonstration participants may be slow, leading to small sample sizes, especially in the treatment arm. Finally, the natural lag associated with claims data shortens the time period available for study prior to submission of this first Report to Congress. After the demonstration projects have been operational for six months or so, we will consult with the project officer regarding the feasibility of including claims analyses in this report. It might, for example, be possible to include preliminary claims analyses for a subset of the demonstration projects.

The baseline analyses of screening and treatment disparities will be summarized for inclusion in this report. This summary will include both the national disparities, and those geographic areas that overlap with the demonstration sites.

2009 Report to Congress. The second Report to Congress will be based on the full set of analyses conducted for this evaluation, including quantitative analyses of screening and treatment completion rates, utilization, Medicare expenditures, cost-effectiveness and cost-utility, quality of life and beneficiary satisfaction, qualitative analyses of provider satisfaction, and findings from the site visits. We recognize that this second report will form the basis for the Secretary's recommendation whether or not to continue the demonstration. Thus, it will be critical that this report clearly answer all of the evaluation questions, but particularly those addressing the demonstration's impacts on racial/ethnic disparities and on Medicare spending.

7.2   Annual and Final Reports

This scope of work for this four-year evaluation originally included three annual reports, with a final report at the end of the project. Given the delays in federal approval of the demonstration sites and thus implementation of the demonstration, the evaluation is now scheduled to run for 57 months, or 9 months longer than planned, ending in June 2010. As a result, we propose to shift the due dates for the "annual" reports in order that they are spread more evenly across the life of the project. In addition, the third annual report will be dropped in order to help fund additional analyses and site visits. The first report on this project will be due in draft in December 2006, with a final version completed by February 2007. The second report will be submitted in draft in February 2008, with a final version in April 2008. The draft final report will be submitted two months before the end of the project as originally scheduled; this date will now be April 2010, with the final report submitted in June 2010.

Each report will contain those analyses conducted to date. The first report will contain descriptive information on each demonstration site, including all available information on enrollments to date. This report will also include preliminary analyses of baseline disparities using Medicare claims and SEER data. The third report will include claims analyses for demonstration participants and site visit findings.

SECTION 8
TIMELINE/DELIVERABLE SCHEDULE

This Section describes the revised schedule of deliverables, which is shown in *Figure 4*. A summary of deliverables is as follows:

- The initial round of site visits is scheduled to begin in December 2006.

- The final survey protocol, and the final OMB package, are due December 2007.

- We anticipate the second round of site visits to be conducted in spring 2009 with a final summary due in the summer of 2009.

- The survey will be implemented from October 2008 – February 2009.

- The draft of the first Report to Congress will be due in March 2008, with a final in May 2008. The draft of the second Report to Congress will be due March 2010, with a final due May 2010.

- There will be one annual report due April 2009.

- All site meetings will occur in May 2006, September 2007, October 2008, and October 2009.

- The final report is due September 2010.

- Final data documentation is due September 2010.

- The final meeting will be scheduled for August 2010.

- Monthly teleconferences and progress reports are ongoing. Analysis of Medicare claims and survey data is ongoing until March 2010.

Figure 4
Schedule of Deliverables

Task	Task Description	Due Date
1	Kickoff Meeting	completed
1a	All Site Meetings	5/06,9/07 completed, 10/08,10/09
2	Initial Site Visits	12/06-8/07
	Protocol	completed
	Summary	in process
3	Design Report/Work Plan	completed
4	Survey Instrument	draft 9/07 completed, final 12/07
5	Prepare OMB Clearance Package	draft11/07, final 12/07
7	Monthly Teleconference	ongoing
8	Monthly Progress Reports and Written Summary of Monthly Teleconference	ongoing
9	Analysis of Medicare Claims and Survey Data	ongoing
10	Site Visits	5/09-7/09
	Protocol	draft 3/09, final 4/09
	Report	draft 5/09, final 6/09
11	Survey Data Collection	10/08--2/09
12	Reports to Congress	drafts 3/08 and 3/10, final 5/08 and 5/10
13	Annual Report	draft 2/09, final 4/09
14	Annual Meeting	3/09
15	Final Report	draft 7/10, final 9/10
16	Data and Documentation	draft 7/10, final 9/10
17	Final Meeting	8/10

REFERENCES

Agency for Healthcare Research and Quality. "National Healthcare Disparities Report: Summary." Agency for Healthcare Research and Quality, Rockville, MD, February 2004. http://www.ahrq.gov/qual/nhdr03/nhdrsum03.htm.

American Cancer Society (ACS). "Cancer Facts & Figures 2004." Atlanta, GA: American Cancer Society, 2004.

Andersen, M.R., N. Urban, S. Ramsey, et al. Examining the Cost-Effectiveness of Cancer Screening Promotion. *Cancer*. 2004 Sep 1;101(5 Suppl):1229-38.

Aziz, N.M. and J.H. Rowland. Trends and Advances in Cancer Survivorship Research: Challenge and Opportunity. *Seminars in Radiation Oncology* 13:248-266, 2003.

Bach, P.B., L.D. Cramer, J.L. Warren, et al. Racial Differences in the Treatment of Early stage Lung Cancer. *N Engl J Med*, 341(16):1198-1205, October 14, 1999.

Baldwin, L.M., S.A. Dobie, K. Billingsley, et al. Explaining Black-White Differences in Receipt of Recommended Colon Cancer Treatment. *J Natl Cancer Inst*. 2005 Aug 17;97(16):1211-20.

Bernabei, R., G. Gambassi, K. Lapane, et al. Management of Pain in Elderly Patients with Cancer. SAGE Study Group. Systematic Assessment of Geriatric Drug Use via Epidemiology. *Journal of the American Medical Association*. 279:1877-1882, 1998.

Brandeis University, Schneider Institute for Health Policy. "Cancer Prevention and Treatment Demonstration for Ethnic and Racial Minorities." 2003. Accessed at http://cms.hhs.gov/healthyaging/CancerPrev.pdf on 1 July 2005.

Briggs, A., and A. Gray. Using Cost Effectiveness Information. *BMJ* 2000;320:246-246.

Brown, M.L., G.F. Riley, N. Schussler, et al. Estimating Health Care Costs Related to Cancer Treatment from SEER-Medicare Data. *Med Care* 2002;40[supp]:IV-104-IV-117.

Chang, S., S.R. Long, L. Kutikova, et al. Estimating the Cost of Cancer: Results on the Basis of Claims Data Analyses for Cancer Patients Diagnosed With Seven Types of Cancer During 1999 to 2000. *Journal of Clinical Oncology*, 2004; 22 (17): 3524-3530.

Chapman, R.H., P.W. Stone, E.A. Sandberg, et al. A Comprehensive League Table of Cost-Utility Ratios and a Sub-table of 'Panel-worthy' Studies, *Medical Decision Making* 2000, 20:451-67.

Coker, A.L., X.L. Du, S. Fang, et al. Socioeconomic Status and Cervical Cancer Survival Among Older Women: Findings from the SEER-Medicare Linked Data Cohorts. *Gynecol Oncol*. 2006 Jan 21.

Cooper, G.S., and S.M. Koroukian. Racial Disparities in the Use of and Indications for Colorectal Procedures in Medicare Beneficiaries. *Cancer*. 100(2):418-24, January 15, 2004.

Crane, L.A., Leakey, T.A., Ehrsam, G., Rimer, B.K., Warnecke, R.B. Effectiveness and cost effectiveness of multiple outcalls to promote mammography among low-income women. *Cancer Epidemiol Biomarkers Prev*. 2000 Sep;9(9):923-31.

Cromwell, J., N.T. McCall, J. Burton, et al. Race/Ethnic Disparities in Utilization of Lifesaving Technologies by Medicare Ischemic Heart Disease Beneficiaries, *Medical Care*. 2005 Apr; 43(4): 330-337.

Denzin, N.K. and Y.S. Lincoln. *Handbook of Qualitative Researchers*: London: Sage, 1994.

Dohan, D., and D. Schrag. Using Navigators to Improve Care of Underserved Patients: Current Practices and Approaches. *Cancer*. 2005 Aug 15; 104(4):848-55.

Fishman, P., S. Taplin, D. Meyer, et al. Cost-Effectiveness of Strategies to Enhance Mammography Use. *Eff Clin Pract*. 2000 Sep-Oct; 3(5):213-20.

French, M.T., M.C. Roebuck and A.T. McLellan. Cost Estimation When Time and Resources are Limited: The Brief DATCAP. *Journal of Substance Abuse Treatment*. October 27, 2004, (3):187-93.

French, M.T., L.J. Dunlap, G.A. Zarkin, et al. A Structured Instrument for Estimating the Economic Cost of Drug Abuse Treatment. The Drug Abuse Treatment Cost Analysis Program (DATCAP). *Journal of Substance Abuse Treatment*, September-October 14, 1997, (5):445-55.

Fryback, D.G., and W.F. Lawrence. Dollars May Not Buy as Many QALYs as We Think: A Problem with Defining Quality-of-Life Adjustments. *Med Decis Making*, 1997; 17:276-84.

Gilligan, T., P.S. Wang, R. Levin, et al. Racial Differences in Screening for Prostate Cancer in the Elderly." *Arch Intern Med*. 164(17):1858-6, Sept. 27, 2004.

Gold, M.R., J.E. Siegel, L.B. Russell, et al. *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press, 2003.

Haddix, A.C., S.M. Teutsch, P.A. Corso, et al. *A Guide to Decision Analysis and Economic Evaluation*. New York: Oxford University Press, 2003.

Harper S, Lynch J. *Methods for Measuring Cancer Disparities: Using Data Relevant to Healthy People 2010 Cancer-Related Objectives*. NCI Cancer Surveillance Monograph Series, Number 6. Bethesda, MD: National Cancer Institute, 2005. NIH Publication No. 05-5777.

Hede K. Agencies Look to Patient Navigators to Reduce Cancer Care Disparities. *J Natl Cancer Inst*. 2006 Feb 1;98(3):157-9.

Holden, D.J., K.S. Moore, and J.L. Holliday. Health Education for a Breast and Cervical Cancer Screening Program: Using the Ecological Model to Assess Local Initiatives. *Health Educ. Res*. 13(2):293-9, 1998 June.

Institute of Medicine. *Promoting Health: Intervention Strategies from Social and Behavioral Research*. Washington, D.C., National Academy Press, 2000.

Krueger, R. and M.A. Casey. *Focus Group Interviews: A Practical Guide for Applied Research.* 3rd ed. Thousand Oaks, CA: Sage, 2000.

Lathan, C.S., B.A. Neville, and C.C. Earle. The Effect of Race on Invasive Staging and Surgery in Non-Small-Cell Lung Cancer. *J Clin Oncol*. 2006 Jan 20;24(3):413-8.

Lynch, F.L., E.P. Whitlock, B.G. Valanis, et al. Cost-Effectiveness of a Tailored Intervention to Increase Screening in HMO Women Overdue for Pap Test and Mammography Services. *Prev Med.*, 2004 Apr;38(4):403-11.

Mauskopf, J., F, Rutten, and W. Schonfeld. Cost-Effectiveness League Tables: Valuable Guidance for Decision Makers? *Pharmacoeconomics*, 2003;21(14):991-1000.

Miles, M. and A.M. Huberman. *Qualitative Data Analysis: An Expanded Sourcebook*. 2nd ed. Thousand Oaks, CA: Sage, 1994.

National Institute of Health, National Cancer Institute (NIH/NCI). "Voices of a Broken System: Real People, Real Problems. President's Cancer Panel." Report of the Chairman, 2000-2001. Bethesda, MD: National Institute of Health, National Cancer Institute, 2001. <http://156.40.135.142:8080/webisodes/pcpvideo/voices_files/PDFfiles/ PCPbook.pdf>.

Ness, R.M., A.M. Holmes, R. Klein, et al. Utility Valuations for Outcome States of Colorectal Cancer. *Am J Gastroenterol*, 1999;94:1650-1657.

Ness, R.M., A.M. Holmes, R. Klein, et al. Outcome States of Colorectal Cancer: Identification and Description Using Patient Focus Groups. *Am J Gastroenterol*, 1998;93:1491-97.

Ogedegbe, G, A.N. Cassells, C.M. Robinson, et al. Perceptions of Barriers and Facilitators of Cancer Early Detection among Low-Income Minority Women in Community Health Centers. *Journal of the National Medical Association*, 97(2):162-170, 2005.

Patton, M.Q. *Designing Qualitative Studies, In Qualitative Research & Evaluation Research*, 3rd ed. Thousand Oaks, CA: Sage; 207-258, 2002.

Salome, H.J., and M.T. French. Using Cost and Financing Instruments for Economic Evaluation of Substance Abuse Treatment Services. *Recent Dev Alcohol*. 2001;15:253-69.

Saywell, R.M. Jr., V.L. Champion , C.S. Skinner, et al. A Cost-Effectiveness Comparison of Three Tailored Interventions to Increase Mammography Screening. *J Womens Health* (Larchmt), 2004 Oct;13(8):909-18.

Subramanian, S., D. Ekwueme, J. Gardner, et al. Theory to Practice: Framework for the Economic Evaluation of the National Breast and Cervical Cancer Screening Program (NBCCEDP). Presented at the International Health Economics Association (iHEA), Barcelona, Spain, 2005.

U.S. Department of Health and Human Services (USDHHS), Public Health Service, National Cancer Institute. 125th National Cancer Advisory Board, Summary of Meeting. Bethesda, MD: National Institutes of Health, February 11-12, 2003. <http://deainfo.nci.nih.gov/ADVISORY/ncab/125_0203/11feb03mins.pdf>.

U.S. Department of Health and Human Services (USDHHS). Making Cancer Health Disparities History. Report of the Trans-HHS Cancer Health Disparities Progress Review Group, March 2004. <http://www.chdprg.omhrc.gov/pdf/chdprg.pdf>.

Wolff, M., T. Bates, B. Beck, et al. Cancer Prevention in Underserved African American Communities: Barriers and Effective Strategies—Review of the Literature. *WMJ*, 102(5):36-40. 2003.

Zarkin, G.A., R.C. Lindrooth, B. Demiralp, et al. The Cost and Cost-Effectiveness of an Enhanced Intervention for People with Substance Abuse Problems at Risk for HIV. *Health Serv Res.*, 2001 Jun;36(2):335-55.