

December 2003

Demonstration to Maintain Independence and Employment Evaluation Strategy

Final Report

Prepared for

Arthur Meltzer, Ph.D.
Center for Medicare & Medicaid Services
7500 Security Boulevard, C3-19-07
Baltimore, MD 21244-1850

Prepared by

Susan G. Haber, Sc.D.
RTI International
Health, Social, and Economics Research
411 Waverley Oaks Road, Suite 330
Waltham, MA 02452-8414

RTI Project Number 08518.002

DEMONSTRATION TO MAINTAIN INDEPENDENCE AND EMPLOYMENT
EVALUATION STRATEGY

FINAL REPORT

by Susan G. Haber, Sc.D., Project Director,

Scientific Reviewer: Janet B. Mitchell, Ph.D.

Federal Project Officer: Arthur Meltzer, Ph.D.

RTI International*

CMS Contract No. 500-00-0046, T.O. #2

December 2003

This project was funded by the Centers for Medicare & Medicaid Services under contract no. 500-00-0046, T.O. #2. The statements contained in this report are solely those of the authors and do not necessarily reflect the views or policies of the Centers for Medicare & Medicaid Services. RTI assumes responsibility for the accuracy and completeness of the information contained in this report.

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

CONTENTS

Introduction.....1

Evaluation Issues1

Evaluation Design.....3

Comparison Group.....5

Sample Size.....6

Data Sources6

DEMONSTRATION TO MAINTAIN INDEPENDENCE AND EMPLOYMENT EVALUATION STRATEGY

Introduction

The Demonstration to Maintain Independent and Employment (DMIE) is intended to test whether providing Medicaid coverage to working people with potentially disabling conditions helps them remain in the workforce, improves their health status and quality of life, and reduces their reliance on SSDI, SSI, and other forms of public assistance. The DMIE legislation requires an independent evaluation of each funded demonstration program in order to determine whether these goals have been achieved. To meet this requirement, each applicant is expected to describe how it would conduct an independent evaluation of its DMIE program. CMS encourages states to partner with an independent organization to conduct the evaluation. The costs of the evaluation are eligible for federal reimbursement.

Any evaluation must be tailored to the specific populations and services covered under a particular demonstration program. This document provides an overview of 1) those areas related to the evaluation that must be addressed by the applicant in its proposal and 2) provides guidance on the expectations for demonstration design as they relate to the evaluation design. The degree to which an applicant addresses the manner in which it intends to design its demonstration and conduct its evaluation to comply with these guidelines will be considered in determining the success of its application.

Evaluation Issues

The evaluation shall consist of two main parts: (1) an impact evaluation and (2) a process evaluation. Following is a description of the types of issues that should be addressed by the process and impact evaluations. These should be seen as illustrative examples and the specific issues addressed by the evaluation should be tailored to the design of each program.

Process Evaluation. The process evaluation should address program implementation, operation, and service delivery as compared to the project plan or proposal. Issues addressed by the process evaluation include, but are not limited to, the following:

- Outreach
 - How were potentially eligible people identified and recruited into the program?
 - What types of outreach activities were undertaken?
 - Were community members (e.g., employers, service organizations, people with disabling conditions, providers, agencies, advocates) aware of the demonstration? Were they involved in outreach?
- Enrollment

- Did the program reach its enrollment goals?
- What were the barriers to enrolling people in the program and to keeping them enrolled?
- How did eligibility determination and redetermination procedures work? What were the procedures for documenting that applicants met the employment, medical, and other eligibility requirements?
- What were the main reasons that applicants were found ineligible for the program?
- For those enrollees that subsequently lost eligibility, what were the main reasons for disenrollment?
- What was the impact of any premiums on program enrollment?
- What types of conditions did program enrollees have and what was the level of severity?
- Service Delivery
 - What was the program benefit package? Were additional services beyond those covered under traditional Medicaid provided to support employment?
 - Did the demonstration provide the range of services required to achieve its goals or were additional types of services needed (for example, case management, employment counseling and support, personal assistance, adaptive technology)?
 - Did enrollees face any barriers to accessing services?
 - What was the impact of any copayments or deductibles on service use?

Impact Evaluation. The impact evaluation should address program outcomes in five main areas: (1) program enrollment; (2) service use; (3) employment; (4) health status; and (5) receipt of cash benefits and other forms of public assistance. Following are illustrative examples of outcomes that could be addressed in each of these areas.

- Program enrollment
 - How many people were enrolled in the demonstration?
 - For how long are beneficiaries enrolled?
 - Do they remain enrolled continuously or is there churning in enrollment?
 - Do demonstration enrollees transition to other types of Medicaid eligibility?
- Service use
 - What types and quantities of Medicaid and other health and vocational services are used by enrollees?

- What is the cost of these services?
- Does enrolling in the demonstration improve access to care?
- How does service use by demonstration enrollees compare to that of people with the same conditions who become disabled and enroll in Medicaid?
- Employment
 - What was the impact of the demonstration on loss or maintenance of employment? What were the reasons for job loss (general economic conditions vs. disability)?
 - What was the impact of the demonstration on the level of employment (hours per week, weeks per year)?
 - What was the impact of the demonstration on earnings? On overall income?
 - What was the impact of the demonstration on job changes (changes in type of job, changes in employer)?
- Health status and quality of life
 - Did the demonstration improve enrollees' health status or slow its deterioration?
 - Did the demonstration reduce limitations on enrollees' physical or other daily activities?
- Receipt of cash benefits and other public assistance.
 - What was the impact of the demonstration on the likelihood that enrollees will become eligible for SSDI?
 - What was the impact of the demonstration on the likelihood that enrollees will become eligible for SSI?
 - What was the impact of the demonstration on the likelihood that enrollees will use other forms of public assistance (e.g., TANF, general assistance, Food Stamps, Unemployment Insurance)?

Evaluation Design

In order to determine the impact of the demonstrations on key program outcomes, the evaluation must compare the outcomes for people enrolled in the demonstration to the outcomes that would have been expected in the absence of the demonstration. Following is a description of alternative evaluation designs to achieve this goal, as well as their advantages and disadvantages.

Pre-post design. Under a pre-post design, outcomes for demonstration enrollees following program enrollment are compared to outcomes for this population prior to enrollment. This design requires that the evaluator have access to outcome measures

prior to and following program enrollment. To the extent that they are not available from administrative data sources, collecting pre-enrollment data may be challenging. This can be addressed by collecting baseline data on pre-enrollment experiences of enrollees at the time of enrollment. The primary weakness of this evaluation design is that there may be changes in outcomes over time that are due to factors other than demonstration enrollment. For example, employment outcomes may be affected by general economic conditions. In addition, the population targeted by the demonstrations (people with potentially disabling conditions) is likely to experience changes in health status over time independent of enrollment in the demonstration. Health outcomes also may be affected by the introduction of new treatment modalities during the course of the demonstration period. As a result, it may be difficult to separate the impact of the demonstration from other potential confounding factors in a simple pre-post design.

Post-enrollment comparison group design. Under a post-enrollment comparison group design, outcomes for demonstration enrollees in the post-enrollment period are compared to outcomes during the same time period in a comparison group that is similar to program enrollees except for their enrollment in the demonstration. (Identification of comparison groups is discussed in the following section.) This design assumes that outcomes in the comparison group reflect the outcomes that would have been expected in the enrollee group in the absence of the demonstration. Unlike a pre-post study, the comparison group design is not compromised by the possibility that outcomes are affected by temporal changes other than the implementation of the demonstration and it does not require collection of pre-enrollment data. However, it can be difficult to identify a comparison population that is similar enough to the enrollee population to provide an adequate control. While statistical analyses can control for some differences between the demonstration and comparison groups, there may remain unmeasured differences between the groups that may affect outcomes and confound evaluation of the demonstration's impact using a simple comparison group design.

Difference-in-difference design. The difference-in-difference model addresses the limitations of the pre-post and post-enrollment comparison group designs by combining these strategies. Under the difference-in-difference design, pre-post changes in the enrollee group are compared to changes in a comparison group over the same period of time. The difference between these groups in the change over time is attributed to the effect of enrolling in the demonstration. This design requires pre- and post-enrollment period data for both demonstration enrollees and the comparison group. Under this model, the pre-post comparison allows each group to act as its own control, addressing the main limitation of the comparison group design. Incorporating a comparison group provides a control for changes over time expected in the absence of the demonstration, the main limitation of the simple pre-post design. The key assumption of the difference-in-difference design is that the *change* in outcomes for the comparison group is similar to what would have been expected in the enrollee group in the absence of the demonstration. To the extent that this assumption is violated, the estimated effect of the demonstration will be biased. This might occur if there are changes over time in external factors unrelated to the demonstration that affect the comparison and demonstration enrollee groups differently. In general, however, the difference-in-

difference design is considered a stronger evaluation strategy than either a pre-post or post-enrollment comparison group design.

Comparison Group

The comparison group should be drawn from a population that is matched as closely as possible to the population eligible for the demonstration. However, the comparison group should not be drawn from people who are eligible for the demonstration, but have not enrolled. It is likely that there are systematic differences between eligible non-enrollees and demonstration enrollees that may affect the outcomes of primary interest. To the extent that these differences cannot be measured and controlled for in statistical analyses, they may bias the measurement of program impacts.¹

One strategy is to draw the comparison group from geographic areas where the demonstration is not operating. In states that have targeted their demonstration to limited geographic regions, the control group can be drawn from other parts of the state that are similar to the area where the demonstration is offered. In the case of statewide demonstrations, the control group may be drawn from another state.

Comparison group members should meet the eligibility requirements for the demonstration if it were offered in the area where they live. These include age, income, employment, presence of the targeted potentially disabling conditions, and any other eligibility requirements for enrollment in the demonstration.

It may also be desirable to enroll comparison group members who are matched with enrollees on additional key characteristics expected to affect outcomes. These might include demographic characteristics (for example, gender, race, or age group) and health status.² Depending on how the comparison group is enrolled into the study, it may not be feasible to match comparison group members directly with enrollees. In this case, it may be possible to control for the impact of these additional characteristics in statistical analyses assuming data on these characteristics are collected. Such an approach may permit the assessment of the independent role of each of these characteristics on the outcome variable(s).

-
- 1 If a state elects to randomly assign people who apply to the program to either the demonstration group or the comparison group, there is no selection bias associated with enrollment in the program. While this is the most “ideal” strategy from the perspective of evaluation design, it may be difficult to implement.
 - 2 As an alternative to direct matching on a set of characteristics, propensity scores can be used to identify a comparison group whose characteristics are balanced with those of the demonstration group, although specific matched study members are not alike in every way. With propensity score matching, observable characteristics are used to predict the probability of enrolling in the demonstration (or the propensity score) for both enrollees and potential members of the comparison group. Each enrollee is matched to the potential comparison group member with the closest propensity score within a defined range. This approach may not be feasible unless the comparison group is identified from administrative data and there is a relatively large pool of potential comparison group members.

Sample Size

The sample size required for the evaluation depends on a number of factors. These include:

- the magnitude of the change in an outcome that the study is expected to be able to detect;
- the variability of the outcome of interest;
- the desired level of statistical significance; and
- the desired level of statistical power.

Statistical significance of $p < .05$ and power of .80 are commonly used levels. To determine the required sample size, evaluators will need to make assumptions about the variability of outcomes and the magnitude of the change that they will be able to detect.

Data Sources

Evaluations of these demonstrations may draw on a variety of data sources. Following are examples of the types of administrative data that may be used.

- Medicaid claims and eligibility data
- SSI data
- SSDI data
- Data on other forms of public assistance
- Unemployment insurance system data

It is likely that data on some program outcomes will not be available from administrative data sources and the evaluation will need to conduct primary data collection. For example, information on health status and quality of life is not available from other data sources. Data on health care service use for comparison group members and for enrollees prior to enrollment are not available from administrative data. It also may not be possible to evaluate certain impacts on employment or use of public assistance services using administrative data. Primary data might be collected through a survey of enrollees and comparison group members, from employers, or from health care providers.

In order to observe longer run program impacts, data for demonstration group members should be collected through the entire evaluation period, regardless of whether the individual remains enrolled in the demonstration. If there is a matched comparison group, data collection for these individuals should follow a comparable time frame. Evaluations that incorporate a pre-post component will also require collection of baseline data. Ideally, 1-2 years of baseline data would be available. To the extent they are not available from administrative data sources and must be based on self-report, reliable baseline data may only be available for shorter time periods.

Applicants may wish to supplement their quantitative evaluation strategy with a qualitative evaluation component involving solicitation of the opinions of program staff, demonstration beneficiaries, and others on such issues as experiences, expectations and perceptions of the program. These opinions may be collected through focus groups or through structured or semi-structured interviews.