



**U.S. Department of Health and Human Services
Assistant Secretary for Planning and Evaluation
Office of Disability, Aging and Long-Term Care Policy**

EVALUATION DESIGN OPTIONS FOR THE LONG-TERM CARE REGISTERED APPRENTICESHIP PROGRAM

September 2011

Office of the Assistant Secretary for Planning and Evaluation

The Office of the Assistant Secretary for Planning and Evaluation (ASPE) is the principal advisor to the Secretary of the Department of Health and Human Services (HHS) on policy development issues, and is responsible for major activities in the areas of legislative and budget development, strategic planning, policy research and evaluation, and economic analysis.

ASPE develops or reviews issues from the viewpoint of the Secretary, providing a perspective that is broader in scope than the specific focus of the various operating agencies. ASPE also works closely with the HHS operating divisions. It assists these agencies in developing policies, and planning policy research, evaluation and data collection within broad HHS and administration initiatives. ASPE often serves a coordinating role for crosscutting policy and administrative activities.

ASPE plans and conducts evaluations and research--both in-house and through support of projects by external researchers--of current and proposed programs and topics of particular interest to the Secretary, the Administration and the Congress.

Office of Disability, Aging and Long-Term Care Policy

The Office of Disability, Aging and Long-Term Care Policy (DALTCP), within ASPE, is responsible for the development, coordination, analysis, research and evaluation of HHS policies and programs which support the independence, health and long-term care of persons with disabilities--children, working aging adults, and older persons. DALTCP is also responsible for policy coordination and research to promote the economic and social well-being of the elderly.

In particular, DALTCP addresses policies concerning: nursing home and community-based services, informal caregiving, the integration of acute and long-term care, Medicare post-acute services and home care, managed care for people with disabilities, long-term rehabilitation services, children's disability, and linkages between employment and health policies. These activities are carried out through policy planning, policy and program analysis, regulatory reviews, formulation of legislative proposals, policy research, evaluation and data planning.

This report was prepared through a contract #HHSP23337001T funded by an Interagency Agreement between HHS's ASPE/DALTCP and the U.S. Department of Labor. For additional information about this subject, you can visit the DALTCP home page at http://aspe.hhs.gov/_/office_specific/daltcp.cfm or contact the ASPE Project Officer, Marie Squillace, at HHS/ASPE/DALTCP, Room 424E, H.H. Humphrey Building, 200 Independence Avenue, S.W., Washington, D.C. 20201. Her e-mail address is: Marie.Squillace@hhs.gov.

EVALUATION DESIGN OPTIONS FOR THE LONG-TERM CARE REGISTERED APPRENTICESHIP PROGRAM

Joshua M. Wiener, Ph.D.
Wayne L. Anderson, Ph.D.
RTI International

Daniel Kuehn, M.P.P.
Robert Lerman, Ph.D.
Urban Institute

September 2011

Prepared for
Office of Disability, Aging and Long-Term Care Policy
Office of the Assistant Secretary for Planning and Evaluation
U.S. Department of Health and Human Services
Contract #HHSP23337001T

The project has been funded in partnership between the U.S. Department of Health and Human Services (HHS), Office of the Assistant Secretary for Planning and Evaluation and the U.S. Department of Labor (DOL), Employment and Training Administration, Office of Apprenticeship under Task Order No. HHSP23337001T. The contents of this publication do not necessarily reflect the views or policies of HHS or DOL, nor does mention of trade names, commercial products, or organizations imply endorsement of the same by the U.S. government. RTI International is a nonprofit, nonpartisan research institute headquartered in Research Triangle Park, North Carolina. The views expressed are those of the authors and should not be attributed to RTI International, its board of governors, or its funders. The Urban Institute is a nonprofit, nonpartisan Research and educational organization established in Washington, D.C., in 1968. The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	iii
EXECUTIVE SUMMARY	iv
1. INTRODUCTION	1
2. BACKGROUND ON THE LONG-TERM CARE APPRENTICESHIP PROGRAM	3
3. RESEARCH QUESTIONS	6
3.1. Apprentices	6
3.2. Employers	7
4. CHARACTERISTICS OF LTC RAPS RELEVANT TO EVALUATION DESIGNS	8
4.1. Uniformity of the Intervention	8
4.2. Size of the Programs.....	9
4.3. Length of Intervention	10
4.4. Selection Bias and Comparison Groups	10
4.5. Limited Data Available at Sites.....	12
4.6. Limitations of the Intervention	12
5. ASSESSMENT OF A BROAD RANGE OF EVALUATION OPTIONS	14
5.1. Outcome Variables.....	14
5.2. Evaluation Designs to Determine Effects on Apprentices	16
5.3. Evaluation Designs to Determine Effects on Employers	18
5.4. Data Sources	20
5.5. Control or Comparison Groups	21
5.6. Evaluation of Broad Options.....	23
6. DETAILED DESCRIPTION OF FOUR APPROACHES TO EVALUATING THE LTC RAP	24
6.1. LEHD Quasi-Experimental Evaluation Option.....	24
6.2. Survey Option for Apprentices	36
6.3. Focus Groups.....	43
6.4. Cost-Benefit Analysis	47
7. CONCLUSION	53
8. REFERENCES	58

LIST OF EXHIBITS

EXHIBIT ES-1.	Overview of Methods, Data Collection, and Potential Feasibility	vi
EXHIBIT ES-2.	Overview of Main Evaluation Design Options.....	viii
EXHIBIT 1.	Overview of Methods, Data Collection, and Potential Feasibility	14
EXHIBIT 2.	Potential Outcome Variables for Workers and Employers.....	15
EXHIBIT 3.	Possible Research Designs for Evaluation of Effects of LTC RAP on Apprentices, Ranked from Strongest to Weakest in Scientific Rigor.....	16
EXHIBIT 4.	Research Designs for Evaluation of Effects of LTC RAP on Employers, Ranked from Strongest to Weakest in Scientific Rigor	18
EXHIBIT 5.	Summary of LEHD LTC RAP Impact Analyses	27
EXHIBIT 6.	Occupation-Industry Crosswalk.....	27
EXHIBIT 7.	Advantages and Disadvantages of LEHD Impact Analyses	29
EXHIBIT 8.	Treatment and Comparison Groups for LEHD Analyses.....	31
EXHIBIT 9.	Example of Question Used to Assess Productivity Gains.....	51
EXHIBIT 10.	Overview of Potential Evaluation Design Options to Evaluate the LTC RAP	54

ACKNOWLEDGMENTS

The authors would like to acknowledge the useful and insightful comments on an earlier draft of this report provided by Professor Robert Glover of the University of Texas at Austin and Dr. Robyn Stone of LeadingAge. Their expertise added greatly to the report. We also thank our Co-Project Officers, Marie Squillace in the Office of the Assistant Secretary for Planning and Evaluation of the U.S. Department of Health and Human Services and Laura Ginsburg in the Office of Apprenticeship at the Employment and Training Administration of the U.S. Department of Labor for the invaluable insight and guidance they provided throughout the study.

EXECUTIVE SUMMARY

The United States faces a critical need for high-quality long-term care workers. The demand for long-term care services is projected to roughly double between 2000 and 2030 as the population ages (Johnson, Toomey, and Wiener, 2007). The U.S. Department of Labor (DOL) projects that home health aides and home care personal care assistants will be among the fastest growing occupations between 2008 and 2018 (DOL, 2011).

Apprenticeship is a well-established strategy for training workers by combining classroom and experiential learning, and placing workers into careers that offer the opportunity for advancement. Best known for training occupations like plumbers and electricians, the apprenticeship model is now being applied to long-term care occupations. By improving the skills of direct care workers, higher wages may be justified by the greater productivity of workers. By restructuring employment in the long-term care industry, apprenticeship may provide a path for career advancement. This report assesses possible research designs to evaluate Long-Term Care Registered Apprenticeship Programs (LTC RAP).

Background

LTC RAPs, registered by DOL Office of Apprenticeship and developed by employers, employer associations and labor-management organizations, provide formal training and work experience for direct care workers in long-term care settings. Since the program's inception in 2003, 119 long-term care employers have offered LTC RAP employment and training to 4,376 long-term care workers (RTI International/Urban Institute analysis of program data, May 2011).

Registered apprenticeship programs are primarily funded by employers with some assistance with start-up funding from government (including DOL) or foundation grants. The required hours of training for all LTC RAPs far exceed what is normally provided by several orders of magnitude. LTC RAPs include four main components. First, on-the-job training (OJT) occurs at a worker's place of employment. Second, related instruction may take place onsite or at technical or community colleges. Related instruction may occur through various modes of instruction (e.g., in-person, web-based, or correspondence courses). Third, mentoring is often a feature of many apprenticeships, occurring sometimes through workers who have completed apprenticeships themselves. Mentors provide on-the-job coaching and help apprentices identify and acquire competencies needed to perform their jobs successfully. Fourth, a clear wage and career progression is a key component of apprenticeship programs. Wage progressions are often tied to the completion of certain occupational competencies,

either in their related instruction, OJT, or both. This advancement opportunity provides an incentive for the apprentices to acquire skills demanded by employers.

Research Questions

In broad terms, research questions about the LTC RAP can be divided into two groups:

- How does the LTC RAP affect apprentices in terms of earnings, job tenure, job satisfaction, and increased competency?
- How does the LTC RAP affect employer sponsors in terms of job turnover, job tenure, improved quality of care, and increased revenue?

In both cases, the comparison is to similar workers and providers who do not participate in or operate LTC RAPs.

Implications of Characteristics of LTC RAPs Relevant to Evaluation Designs

In the evaluation of any program, the particular characteristics of the intervention make it easier or more difficult to design an evaluation. Some of the characteristics of the LTC RAP that affect possible research designs include:

- *Decentralization of design responsibility to individual employers.* Although employers have great flexibility in how they design and administer their LTC RAPs, there appears to be enough uniformity in goals and programs to be able to talk meaningfully about a single LTC RAP program.
- *Size of the LTC RAP program.* As of May 2011, there were 119 LTC RAPs, 954 active apprentices, a total of 1,347 people who have completed an apprenticeship, and overall, a total of 4,376 apprentices who had ever participated in the program, regardless of whether they completed an apprenticeship or not. Most programs are small with just a handful of apprentices. To have a large enough sample to detect statistically significant effects requires a large number of apprentices, probably requiring the entire program rather than a sample.
- *Availability of data.* Based on our site visits, it appears that few programs collect much systematic data on outcomes. Thus, almost all of the data will need to be collected by an evaluation contractor or from administrative databases collected for other purposes.

- *Selection bias and the problem of comparison groups.* A key characteristic of LTC RAPs is that, for most programs, only a small percentage of direct care workers within an employer sponsor are selected to participate. These workers are typically selected because they are the best and most promising workers. Thus, these workers are likely to differ in important ways from other workers of the same age, gender, education and years of work experience, making development of comparison groups more difficult.
- *Sponsors' use of apprentices to improve non-apprentice staff performance and the problem of comparison groups.* One possible strategy to develop a comparison group is to select people working for the same employer who are not apprentices. However, employers visited during our case studies almost always assigned apprentices to act as peer-mentors for other workers. While this is a strength of the program, it means that non-apprentices are not free of the potential impact of the apprenticeship program and are, therefore, problematic as a comparison group.

Assessment of a Broad Range of Evaluation Options

There are many possible research designs for an evaluation of the LTC RAP, with varying costs and degrees of scientific rigor. Most evaluations of job training programs focus solely on the program participants, mainly their gains in employment and earnings. However, since many of the policy motivations for LTC RAPs have to do with improving the performance and quality of care of long-term care providers, the evaluation should address both apprentices and their employers.

EXHIBIT ES-1. Overview of Methods, Data Collection, and Potential Feasibility			
Analysis Methods	Qualitative Analysis	Quantitative Analysis	
		Descriptive analysis (single point in time AND no comparison group)	Multivariate analysis (two points in time OR with a comparison group OR both)
Data Collection	Focus groups of workers Case studies of employers In-depth ethnographic studies and implementation evaluation	Survey of LTC RAP workers or employers across occupation/ organization types LTC RAP administrative data Linking Medicare/ Medicaid claims/ OSCAR data	Survey only Survey and existing secondary data (National Nursing Assistant Survey; National Home and Hospice Care Survey)
Feasibility	Lower cost Lower generalizability	↔	Higher cost Higher generalizability

The overarching approach for most designs is to compare the apprentice and employer performance to what it would have been in the absence of the LTC RAP. **Exhibit ES-1** provides a broad overview of the range of methods, types of data collection, and their relationship to costs and ability to generalize the findings to the total population of LTC RAPs.

Detailed Description of Four Approaches to Evaluating the LTC RAP

After considering a large number of possible options, the RTI/Urban team devised a four-component approach to evaluating the LTC RAP. The four components are: (1) use of the Longitudinal Employer Household Dynamics (LEHD) administrative dataset to compare workers who have participated in the LTC RAP program with workers who have not participated in the LTC RAP; (2) a cross-sectional telephone survey of workers who have ever participated in the LTC RAP and workers who have never participated in the LTC RAP; (3) focus groups with apprentices and focus groups of employers without a comparison group; and (4) a cost-benefit analysis of the LTC RAP from the employer's perspective. With the exception of the cost-benefit analysis, which depends in part on the analyses of the administrative dataset and the telephone survey, each component is separate, but complementary, and could be funded without the others. Thus, government decision makers can mix-and-match the approaches as they see fit; they can decide to fund any one component or all pieces or any combination. **Exhibit ES-2** summarizes the four components and their advantages and disadvantages. The estimated cost for all four components is \$985,000 in 2011 dollars.

The first design would use the LEHD administrative database to assess the effect of LTC RAPs on increased apprentice earnings and job tenure, and on the worker turnover rate at the employer level. The LEHD is a Census Bureau database that includes state-level Unemployment Insurance administrative information on employment and earnings merged with certain other Census data. The biggest challenge for this design is using the limited variables available in the Unemployment Insurance data to construct a truly comparable comparison group.

In other studies of job training programs, prior earnings are used to proxy many personal characteristics, but wages (although not hours) are highly constrained in long-term care. In addition, the dataset can identify low-wage workers in other long-term care organizations, but cannot separate direct care workers from other low-wage workers in long-term care organizations (e.g., housekeeping and dietary workers in nursing homes and assisted living facilities). Still, this option provides potentially the most viable design to credibly address the most important research questions facing the industry.

EXHIBIT ES-2. Overview of Main Evaluation Design Options

Option	Advantages	Disadvantages
<p>Analysis of LEHD, comparing all apprentices with matched sample comparison group \$285,000 27 months</p>	<ul style="list-style-type: none"> • Uses data on all apprentices, regardless of when they started and whether they completed the program • Captures duration with the firm before, during, and after apprenticeship • Addresses major issues of earnings and job tenure and continued employment in the industry • Dataset likely to include very high percentage of people ever participating in LTC RAPs • Easy access to a large supply of low-earning people working for non-apprentice long-term care providers for comparison group • No new data collection required; no Office of Management and Budget review required 	<ul style="list-style-type: none"> • Limited data on which to match apprentices and comparison group, leaving possibility of uncontrolled for selection bias • No data from perspective of apprentices on outcomes such as job satisfaction • No data from perspective of employers, except for duration of apprentices within the firm • Low-wage workers in comparison group will include housekeepers and dietary staff as well as direct care workers
<p>One-time cross-sectional survey of apprentices and matched comparison group \$450,000 30 months</p>	<ul style="list-style-type: none"> • Addresses more subjective outcomes, such as job satisfaction and relationship with supervisor • Provides more detailed data on apprentices • Possible to more completely control for selection bias 	<ul style="list-style-type: none"> • As cross-sectional design, only able to analyze “association” rather than causation • Comparison group facilities/agencies may be reluctant to provide contact information about apprentices • Correction for selection bias can only made after initial contact since providers unlikely or unable to provide detailed information on workers, raising costs • Only able to include apprentices who have stayed with employer that trained them; apprentices that left employer or field lost to analysis • Less consensus on measurement of “softer” outcomes • More expensive than other options

EXHIBIT ES-2 (continued)		
Option	Advantages	Disadvantages
Focus groups of apprentices and of employers \$150,000 14 months	<ul style="list-style-type: none"> • Low-cost option • Provides information on views of apprentices • Can provide detailed suggestions from participants for improving LTC RAPs 	<ul style="list-style-type: none"> • Qualitative data cannot be used to determine effectiveness of intervention • Representativeness of views expressed cannot be directly assessed • Views by apprentices and providers provided cannot be easily summarized or quantified • Comparisons cannot be made to workers who did not participate in LTC RAP
Cost-benefit analysis \$100,000 14 months	<ul style="list-style-type: none"> • Attempts to measure whether benefits to employer exceeds the costs, which is key for establishing business case for program • Measures changes in turnover related to LTC RAPs • Consistent with approaches used in other studies of apprenticeship costs and benefits • Low-cost data collection 	<ul style="list-style-type: none"> • Measurement of relative productivity of apprentices is not straight forward • Employer estimates may be biased as some try to justify their investments

The second design option is a cross-sectional, one-time telephone survey of apprentices and a comparison group of non-apprentices to determine the effects of apprenticeship on job satisfaction, intent to leave one's job, relations with supervisors and other staff, and other factors that only workers can address. The survey findings could demonstrate an association between the apprenticeship program and outcomes, but causality could not be attributed to the LTC RAP because there are no measures of change over time. Moreover, among direct care staff that stay in their jobs, previous studies have found high rates of job satisfaction, suggesting either that existing measures are not very sensitive or that there is not that much room for improvement among workers who stay in their jobs (Bishop et al., 2009).

The third design option would provide a much more detailed understanding of apprentice and employer opinions about how apprenticeship works. Eight focus groups would be conducted among apprentices at eight different employers, and two focus groups would be conducted among management of employer sponsors. The apprentice focus groups would be held in the general geographic area of the employer, but not at the employer's location; the employer focus groups would be held at national provider association meetings. These focus groups would provide a rich understanding of the value of apprenticeships over traditional training and how employers implement their LTC RAPs, but it could not provide any quantitative estimates of the impact of LTC RAPs.

A fourth evaluation design focuses on the employer-level benefits and costs of the LTC RAPs. Benefits, measured as the increased productivity achieved by the LTC RAPs, and a range of implementation costs would be gathered through an Internet survey process among a selected group of employers. Costs would include supervision time, apprentice time lost to regular work, and whatever curriculum development that the facility or agency does. Data from the LEHD analysis would also be used to determine benefits. The greatest challenge for this design is the lack of data for employers to accurately assess the improvement in performance and productivity due to the LTC RAPs. This design would explicitly address questions related to the business case for employers.

In considering these alternatives, the Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services and DOL must answer two major questions. First, can the LTC RAP be a strong enough intervention to yield net benefits at the apprentice or employer-level? Is it plausible to expect gains in increased wages, job tenure, job satisfaction, commitment to the industry, productivity and quality of care and decreased turnover? In other words, can the LTC RAP approach be implemented on a large enough scale that it can possibly improve outcomes for consumers, workers, employers, clients and funders for a large number of apprentices and employers? Second, can the research designs presented here or other possible designs produce results that can withstand critical scrutiny from researchers and policymakers? In other words, will the evaluation provide methodologically defensible results that justify the cost of the evaluation?

1. INTRODUCTION

The United States faces a critical need for high-quality long-term care workers. The demand for long-term care services is projected to roughly double between 2000 and 2030 as the population ages (Johnson, Toomey, and Wiener, 2007). The U.S. Department of Labor (DOL) projects that home health aides (HHAs) and home care personal care assistants will be among the fastest growing occupations between 2008 and 2018 (DOL, 2011).

Long-term care workers include certified nursing assistants (CNAs), who work in nursing homes; HHAs, who work for home health agencies; health support specialists (HSSs), who work in residential care facilities, and direct support specialists (DSSs), who work in group homes or facilities for persons with intellectual disabilities. These direct care workers assist people with disabilities with daily activities such as bathing, eating, shopping, and housecleaning as in various types of group residential settings as well as in private dwellings.

Low wages, few fringe benefits, minimal levels of training, and the lack of a career ladder contribute to chronic workforce shortages (Stone and Wiener, 2001). Residential care providers and nursing homes report high turnover rates ranging from 40%-70% (National Center for Assisted Living, 2010; American Health Care Association, 2010). Low compensation and few options for advancement result in weak incentives for workers to enter or remain in the long-term care field (Khatutsky, Wiener, Anderson, and Squillace, 2011). Employers are also often dependent on tight Medicaid reimbursement rates, further constraining them from raising wages to attract new workers.

For some direct care workers, federal and state regulations require some minimal training. A federal standard of at least 75 hours of training applies for CNAs in skilled nursing facilities and HHAs in home health agencies. There are no federal (or sometimes even state) training requirements for personal assistant service workers, personal care attendants, aides in assisted living facilities, or direct care workers serving people with intellectual disabilities. To help improve recruitment and retention of direct care workers and to improve the quality of care, the Institute of Medicine (2008) recommended increased training requirements and career development for all direct care workers to address the workforce shortage and to improve the quality of care in long-term care settings.

Apprenticeship is a well-established strategy for training workers by combining classroom and experiential learning, and placing workers into careers that offer the opportunity for advancement. Best known for training occupations like plumbers and electricians, the apprenticeship model is now being applied to long-term care occupations. By improving the skills of direct care workers, higher wages can be

justified by the greater productivity of workers. By restructuring employment in the long-term care industry, apprenticeship can provide a path for career advancement. Long-Term Care Registered Apprenticeship Programs (LTC RAP) is an initiative to expand the apprenticeship concept to long-term care workers. DOL's Committee on Apprenticeship is committed to expanding apprenticeship into emerging industries, including the long-term care sector. DOL also has a goal of expanding employment options for women and the LTC RAP is one vehicle to achieve that goal.

Apprenticeships in long-term care are more common in other countries. In the United Kingdom, health and social care apprenticeships, such as health care assistants, are increasingly common. For example, Barchester Health Care, a large British company operating 200 nursing care homes with over 10,000 residents, uses apprenticeships extensively for long-term care workers (Mansfield-Loynes, 2011) and reports that apprenticeships reduce worker turnover significantly (personal communication, Terry Tucker, Director of Learning and Development, Barchester Healthcare, July 28, 2011). Long-term care apprenticeships also exist in Australia and Germany. The fact that these types of apprenticeships are under way in a number of countries suggests that apprenticeship training for long-term care workers is feasible.

This report is the final deliverable of a joint project sponsored by the Office of the Assistant Secretary for Planning and Evaluation (ASPE), U.S. Department of Health and Human Services (HHS) and DOL to assess the feasibility of conducting a rigorous evaluation of the LTC RAP. This analysis evaluates possible research designs to evaluate the LTC RAP administered by DOL. Section 2 provides background on the LTC RAP. Section 3 discusses the key research questions that should be addressed by an evaluation of the LTC RAP. Section 4 discusses some characteristics of the program that are particularly important in considering evaluation research designs. Section 5 discusses a wide range of possible research designs, briefly assessing their advantages and disadvantages. Section 6 describes in detail four complementary research designs that could be used to evaluate the LTC RAP. Section 7 concludes with an analysis of the main evaluation designs to evaluate the LTC RAP. This report builds on two previous papers on the LTC RAP conducted by RTI International and the Urban Institute under this contract -- one paper analyzes administrative data on the LTC RAP, which is maintained by the DOL (Anderson et al., 2010) and the other paper reports on site visits to five LTC RAPs (Kuehn et al., 2011).

2. BACKGROUND ON THE LONG-TERM CARE APPRENTICESHIP PROGRAM (LTC RAP)

The apprenticeship model is distinguished by its integration of instruction and work; apprentices learn occupational competencies both in formal classroom settings and while working at a job that directly applies and reinforces those competencies. Structuring training in this way provides apprentices with an income during the training and helps assure that the skills they learn are useful to employers. In addition, the work-based learning offered by on-the-job training (OJT) helps apprentices to understand how their classroom instruction is relevant to their work. An essential component of apprenticeship is a clear wage and career progression. Wage progressions are often tied to the completion of certain occupational competencies, either in their classroom instruction, OJT, or both. This advancement opportunity provides an incentive for the apprentices to acquire skills demanded by employers.

Apprenticeship in the United States is highly decentralized with decisions made by individual apprenticeship sponsors regarding curriculum and program structure. Most programs operate within the Registered Apprenticeship system, which is overseen by DOL's Office of Apprenticeship (OA) and state apprenticeship agencies. The OA and the state apprenticeship agencies certify program completion, protect the safety and welfare of apprentices, provide guidance and technical assistance to program sponsors, monitor program equal opportunity plans to prevent discrimination against women and minorities, and promote the expansion of the use of apprenticeship by employers. Only an apprenticeship program registered with the OA or a State Apprenticeship Agency and meeting the minimum requirements for standards of apprenticeship established in 29 CFR 29.5, can receive certification and be recognized across the country. Almost no direct government funds are spent on apprenticeship programs, including LTC RAP.

LTC RAPs, registered by the OA and developed by employers, employer associations and labor-management organizations, provide formal training and work experience for direct care workers in long-term care settings. Since the program's inception in 2003, 119 long-term care employers have offered LTC RAP employment and training to 4,376 apprentices, including all workers regardless of whether they completed the apprenticeship or not (RTI International/Urban Institute analysis of program data, May, 2011).

Registered apprenticeship programs are primarily funded directly by employers with some assistance with start-up funding from government (including DOL) or private organization (e.g., foundation) grants. LTC RAPs include four main components. First, OJT occurs at a worker's place of employment. Second, related instruction takes place either at the work site or at technical or community colleges. Related instruction may occur through various modes of instruction (e.g., in-person, web-based, correspondence course). Third, mentoring is often a feature of many apprenticeships,

occurring sometimes through mentors who have completed apprenticeships themselves. Mentors provide on-the-job coaching and help apprentices identify and acquire competencies needed to perform their jobs successfully. The required hours of training for all of the LTC RAPs far exceeds what is normally provided by several orders of magnitude. For example, Agape Senior's LTC RAP for CNAs in South Carolina requires 2,257 hours to complete, and includes 266 hours of related training instruction, which is more than three times the minimum federal requirement of 75 hours of training. Fourth, a clear wage and career progression is a key component of apprenticeship programs. Wage progressions are often tied to the completion of certain occupational competencies, either in their classroom instruction, OJT, or both. This advancement opportunity provides an incentive for the apprentices to acquire skills demanded by employers. Ideally, completion of the apprenticeship results in a job certification that is portable and meaningful to other employers.

Registered apprenticeships are structured to develop increased job competency over time. Apprenticeships can be either competency-based, time-based, or a hybrid of the two, a decision made by sponsors who can shape OJT and curricula previously developed by the OA to suit employer needs. All programs require apprentices to master a set of competencies, but like most schools, the time-based approach additionally require certain minimum hours of on-the-job and related instruction. Hybrid programs often require minimum time spent in on-the-job or related training.

Current LTC RAPs offer apprenticeships in four major occupations: CNAs, DSSs, HSSs, and HHAs. CNAs, work in nursing homes caring for persons with clinical needs or needing assistance with eating, bathing, and similar activities. The CNA apprenticeship is competency-based and offers two models, one with interim credentials and one without. In the interim credentialed model, apprentices must complete entry-level and advanced level training, which is followed by one or more specialties such as dementia or restorative care. After completion of each level and each specialty, apprentices receive a Certificate of Training. Upon completion of Levels 1, 2 and any specialty from Level 3, apprentices receive a Certificate of Completion of Apprenticeship.

DSSs provide care in group homes for persons with intellectual and developmental disabilities needing monitoring and assistance in daily activities. The DSS apprenticeship is a competency-based model with no interim credentials offered.

HSSs work in assisted living facilities and other residential care facilities providing care for mostly elderly persons needing monitoring and assistance with daily tasks. These residential care facilities usually do not provide the highly skilled clinical care that is provided in nursing homes, thus staff certification requirements for these two settings differ. The HSS apprenticeship currently is a hybrid model (time-based and competency-based) with no interim credentials.

HHAs work in home health and hospice agencies providing services to people living in the community having clinical needs or needing assistance with eating, bathing,

and similar activities. The HHA model is a competency-based apprenticeship offering interim credentials or Certificates of Training when apprentices complete various levels of training within the occupation. The apprenticeship begins with entry-level (Level 1) training, of which at least 16 hours of classroom training must be completed before beginning the supervised practical training component. To receive the Certificate of Completion of Apprenticeship, apprentices need to complete Level 1 and then any two specialties.

Specialty training differs across the three occupations that have such training. CNAs can specialize in dementia care, geriatric care, restorative care, or mentoring. HHAs can specialize in care for people with disabilities, palliative care for patients receiving hospice, care for people with mental illness, dementia care, geriatric care, or mentoring. HSSs can receive specialty training in dining services, environmental services, or as an activity director, certified medication aide, certified nurse's aide, HHA, or rehabilitative aide.

Generally, competency-based apprenticeship programs emphasize skill mastery without requiring a specified time commitment to training, although OJT ranges from 3 to 5 months for each level of training with varying amounts of related instruction. In contrast, time-based apprenticeships mandate occupational competencies apprentices must learn for certification within a fixed amount of time at the end of which participants receive a certificate. Time-based apprenticeships generally provide an extended period of entry-level training lasting a minimum of 2,000 hours, with at least 144 hours of related instruction.

3. RESEARCH QUESTIONS

Whether an evaluation of the LTC RAP is feasible and which design is best depends in part of what questions ASPE/HHS and DOL wish to answer. For example, questions about how apprentices view the LTC RAP can be answered with focus groups or surveys of workers, but cannot be answered with administrative datasets, such as Unemployment Insurance data. Conversely, questions about how the LTC RAP affects earnings can be answered with an administrative dataset, but not with focus groups.

In broad terms, the research questions about the LTC RAP can be divided into two groups:

- How does the LTC RAP affect apprentices, including participants who do not complete the program?
- How does the LTC RAP affect long-term care employer sponsors?

In both cases, the comparison is to workers and providers who do not participate in or operate LTC RAPs. Detailed research questions are presented in Section 6, which describes a comprehensive approach to evaluating the program.

3.1. Apprentices

Most of the research questions on apprentices relate to whether the training improves the income and skills of LTC RAP participants and the workers' commitment to the employer and the field of long-term care. These questions can be asked of both participants who complete the apprenticeship program and those who do not complete the program. Important questions include:

- Does participation in the LTC RAP improve the earnings of apprentices?
- Does the LTC RAP improve the skills and productivity of the apprentices?
- Does the LTC RAP increase job satisfaction and improve relationships with other staff, supervisors, and clients?
- Does the LTC RAP increase job tenure and the likelihood of continuing to work in long-term care?
- Does LTC RAP improve access to continued career path/ladder opportunities?

- Are LTC RAP credentials portable to other employers, health care sectors, and regions?

3.2. Employers

Most of the research questions on employers relate to whether the training improves outcomes for the organization as a whole. Important questions include:

- Does the LTC RAP improve the quality of care/quality of life provided by the organization?
- Does the LTC RAP reduce turnover and, therefore, reduce recruitment and new training costs?
- Does LTC RAP help employers to reduce other costs such as worker compensation and insurance premiums?
- Does the LTC RAP increase the firm's revenues?
- Does the LTC RAP improve the organizational climate and relationships between staff members and management and with clients?

4. CHARACTERISTICS OF LTC RAPS RELEVANT TO EVALUATION DESIGNS

In the evaluation of any program, the particular characteristics of the intervention make it easier or more difficult to design an evaluation. The LTC RAP has numerous characteristics relevant to designing an evaluation including: (1) the uniformity of the intervention; (2) program length; (3) the size and scalability of these programs; (4) availability of data; (5) recruitment and selection of apprentices into the programs; and (6) implications of sponsors' use of apprentices to improve non-apprentice staff performance.

4.1. Uniformity of the Intervention

The design of LTC RAPs is decentralized with individual employers given great discretion over what goals they attempt to achieve and the critical elements of the training program, including the curriculum, length of the apprenticeship, and types of training provided. LTC RAPs include four different occupations, with different training requirements both across occupations and within occupations. A key issue is whether the sites' program goals and the interventions they administer are uniform enough that the program can be evaluated as a whole. Successful evaluations of multi-site programs require that the goals and the activities of the different sites be at least roughly the same; if they are not, then it is not clear what intervention is being evaluated.

In the RTI International/Urban Institute analysis of the administrative data and in our site visits, we found that while there was variation across sites, there was enough uniformity across sites in our judgment to evaluate the program as a whole (Anderson et al., 2010; Kuehn et al., 2011). For example, in all of the sites visited, the goals were to improve the long-term care workforce in order to improve quality of care and to create more attractive jobs for apprentices. These goals help sponsors to meet state certification requirements, reduce errors in caregiving, reduce turnover, and create career opportunities for apprentices.

Moreover, while the length and content of the components of the LTC RAP programs varied, they all included at least the basic structure of the apprenticeship programs -- OJT, related instruction, peer-mentorship, and a wage increase upon the successful completion of the program. While the distinction between routine supervision and OJT was sometimes unclear, all of the training programs visited involved substantially greater levels of training than is typically required by federal and state regulation. Although one site used its LTC RAP for entry-level training of all new employees, most sites used the LTC RAP for advanced training and mentoring of employees who had already received basic training and had leadership or other personal qualities that management wished to develop.

There are, however, substantial differences in the content of the training provided and modalities of delivery across occupations in terms of the skills learned to meet the needs of different service settings and populations. Different settings also have an impact on how some of the training can be administered. For example, while CNAs work in a nursing home where supervisors are readily available, HHAs provide services in the homes of individuals, usually without direct supervision. Moreover, there may be a substantial difference between the program described in the agreement with DOL and the program that is actually implemented. As a result, other observers may judge the programs too heterogeneous to analyze as a whole; with that judgment, separate evaluations would have to be done of each occupation, which would have a major impact on the sample sizes potentially available for analysis. One partial solution to this problem is to use the four occupations as control variables in the multivariate analyses.

4.2. Size of the Programs

LTC RAP is a program characterized by a large number of programs spread across four different occupations with few apprentices in each program (Anderson et al., 2010). As of May 2011, the entire LTC RAP consists of 119 training programs, 954 active apprentices, a total of 1,347 people who have ever completed an apprenticeship, and overall, a total of 4,376 apprentices who had ever participated in the program. Based on national Registered Apprenticeship Partners Information Data System (RAPIDS) data, LTC RAPs have a median size of only six active apprentices. As of May 2011, there were only about seven sites with more than 25 active apprentices. The size of the LTC RAPs visited for the case studies ranged from eight to 183 active apprentices as of May 2011 (Kuehn et al., 2011). Moreover, the active apprentices are divided up among the four occupations; 515 CNAs, 284 DSSs, 107 HSSs, and 48 HHAs. CNAs account for half of the total number of apprentices.

The large number of small programs has several important evaluation design implications. First, in order to obtain a large enough sample size to detect statistically significant results in the outcome variables, such as the annual turnover rate, it will be necessary to include all or at least a large share of LTC RAPs and LTC RAP current and former apprentices. Including all 119 programs will mean collecting the data by mail or telephone surveys rather than in person since it will be too expensive to visit each program.

Second, because of the relatively small number of apprentices within the different occupations, the ability to conduct subgroup analysis is limited. While analysis of the program as a whole will be possible, statistical power analyses suggest that subgroup analysis of occupations will need to be limited to CNAs and DSSs. The number of people in the HHA and HSS LTC RAPs is too small for subgroup analyses.

Third, with some exceptions, within each facility/agency, the number of people who have received training through the apprenticeship program is a small proportion of the

total number of workers for participating long-term care employers. As a result, the impacts on facility/agency-level outcomes, such as turnover and retention and quality of care, may be small because there is not a critical mass of people to affect those organization-level outcomes.

4.3. Length of Intervention

The length of the LTC RAP programs is an important issue for any evaluation because of its implications for the cost of data collection. An evaluation that involves measuring the apprentice at the beginning and the end of the program (and perhaps during the course of the program and afterwards) would need to consider how long it is necessary to follow an entrant to the program. The longer the program, the more difficult it is to gather information, the more likely that some apprentices will not complete the program, and the more likely that apprentices will be lost to follow-up. The programs visited on the site visits have a wide range in time for completion, with the shortest program being 232 hours and the longest program being 3,000 hours (approximately 1.5 years) (Kuehn et al., 2011). The two remaining programs were approximately 2,000 hours, which is a full year.

Length of the intervention is also important because longer evaluations are usually more expensive than shorter ones, particularly if they involve multiple waves of data collection. In order to follow entrants to the LTC RAP, data would have to be collected on new entrants to the program for at least a year, if not longer, to obtain a large enough sample size, and then followed on a flow basis for another year. Because of the small number of people in the LTC RAP, it would take a long period of data collection to build up an adequate sample size, which would be expensive.

Finally, the longer the program, the more likely it will be that apprentices will not complete the program. Some apprentices will find it too difficult to continue; others may leave the employment of the LTC RAP sponsor, either to go to another long-term care provider or to leave the field. As a result, to the extent possible, analyses will need to be done on several groups -- all persons who ever participated in the program, current apprentices, and direct care workers who have completed the apprenticeship program. Sample sizes will dictate how many of these different analyses can be actually conducted.

4.4. Selection Bias and Comparison Groups

The essence of evaluation research is asking whether the outcomes of the intervention group are different from those of some comparable comparison group which did not receive the intervention. Developing comparison groups for LTC RAP apprentices and programs will be challenging because of the selection bias inherent in the way apprentices are chosen for the program. Based on our site visits, most programs have selection criteria for apprenticeships; they are not open to all workers

and they are not randomly chosen. Employees must typically apply or be recommended by a supervisor and are selected based on qualities such as their superior caregiving abilities, intelligence, ambition, and ability to work with clients and other staff. These attributes are not variables in administrative datasets that could be used to construct a comparison group. As a result, apprentices are likely to be different from similar workers of the same age, gender, race, education, and employment history. Thus, comparisons between apprentices and comparison groups that do not control for selection bias may measure outcomes that are the result of differences in the personal characteristics of the workers rather than the impact of the apprenticeship program.

The classic solution to the problem of selection bias is a randomized controlled trial, which randomly assigns all participants either to the intervention or the control group. Thus, people with unmeasured differences are equally likely to be in either the treatment or the control group. However, without random assignment, evaluators must seek other options to distinguish between program effects and effects linked to unmeasured individual differences. One approach is to gather more information about the comparison group and match people in the intervention with people not in the intervention. Arguably, prior earnings could be a proxy for some of the personality characteristics that may be important in the choice of apprentices and could be used to match non-apprentice workers for a comparison group. Gathering information not in administrative databases can be done, but it increases the expense in selecting the comparison group because not all people on which the information is collected will be used in the comparison group. Moreover, this approach does not guarantee that the biasing factor will be identified and measured. An approach that can control for measured differences in individuals in the treatment group and the comparison group is multivariate analysis which can statistically control for many variables, but it cannot control for unmeasured differences in skill level, experience, motivation, and aptitude for service in long-term care.

A further complicating factor is that most programs are designed so that apprentices who complete the program serve as mentors to the non-apprentice staff. While this is strength of the program and builds the business case for LTC RAP, this program design means that non-apprentices in the same facility/agency are not free of the potential impact of the apprenticeship program and are, therefore, inappropriate as a comparison group. To address this issue, an evaluation would need a comparison group outside of the sponsor's organization, or at least another of the sponsor's facilities or agencies not participating in the intervention.

Another issue concerning comparison groups is the difficulty of convincing the comparison group to participate in the evaluation since they are not participating in the LTC RAP. For administrative datasets, obtaining cooperation is not a barrier to the study because the permission of the employers and workers is not required for research purposes. However, developing a comparison group for a survey of employers or apprentices may be difficult. While employers sponsoring LTC RAPs and apprentices presumably have some interest and motivation in participating in a study about the LTC RAP, employers in comparison groups may not be eager to provide information on their

training practices to outside organizations and may be reluctant to release confidential contact information of employees. For their part, direct care workers may see little reason to participate in the survey. Offering survey respondents a modest financial incentive is a common strategy to increase participation.

Finally, ideally, an evaluation would compare apprenticeship with standard training, but with no advanced training. While minimal training is the norm in long-term care, some providers do provide more extensive training that has elements of the LTC RAP. Indeed, some of the providers in the case studies reported that they had enhanced training programs prior to their implementation of the LTC RAP. To the extent that the comparison group offers enhanced training, it will be more difficult to detect the influence of the LTC RAP.

4.5. Limited Data Available at Sites

One possible source of data in some evaluations is administrative data that is routinely collected by the employer sponsors whose programs are being evaluated. LTC RAPs report some data to DOL through RAPIDS. In terms of evaluation, RAPIDS data is useful for identifying the universe of sponsors and apprentices, and some socioeconomic characteristics and Social Security numbers of apprentices and Employer Identification Numbers of employers, but it does not have outcome data and some of the data is not updated on a regular basis. Moreover, during the site visits, providers reported that they collected very limited data on outcomes or costs. Most sites did collect data on wages, benefits, tenure, and turnover, but not in a common format across sites. Thus, data routinely collected by the LTC RAPs is unlikely to provide much needed data for the evaluation. New data will need to be collected for an evaluation.

4.6. Limitations of the Intervention

The LTC RAP represents a more substantial commitment to training than is typical in long-term care. As an approach to training in the long-term care industry, LTC RAPs are still in their early stages and have not diffused fully across the industry. Among the strengths of these apprenticeship programs are the emphasis on mentors and peer-to-peer learning, learning by doing, and the integrated learning through theory and practice. Nonetheless, compared to traditional apprenticeship programs for occupations like plumbers, electricians or carpenters, the LTC RAP faces numerous serious challenges.

- It is a relatively new approach in the long-term care field; as yet, few providers in long-term care have heard of or are knowledgeable about the LTC RAP; as a result, completing the apprenticeship program provides little recognition outside of the sponsoring employer. As a result, the certificates of completion may not have value as a credential respected by other employers for entry into jobs with

higher salaries and access to more responsible positions. If the program were to expand, the apprentice's completion credential might increase in value.

- Employers, often constrained by payments from public programs (principally Medicaid), are unable or unwilling to provide substantial wage increases to apprentices completing the apprenticeship program. In our site visits, apprentices completing the program received wage increases of \$0-\$1.25 per hour. This is consistent with other findings that there is little wage growth with longer job tenure among CNAs working in nursing homes (Wiener, Squillace, Anderson, and Khatutsky, 2009).
- Completing the apprenticeship is not a rung on a well-established career ladder. Few providers have job titles for direct care workers with advanced training, although some states are beginning to recognize career lattices as well as ladders in their regulations and nurse delegation legislation. Moreover, apprenticeship experience and training as CNAs, HHAs, DSSs and HSSs does not typically allow them to move to higher level clinical or management positions. In almost all cases, moving up in the organization requires obtaining additional formal education (e.g., becoming a licensed practical nurse requires going back to school).

5. ASSESSMENT OF A BROAD RANGE OF EVALUATION OPTIONS

There are many possible research designs for an evaluation of the LTC RAP, with varying costs and degrees of scientific rigor. Most evaluations of job training programs focus solely on the program participants, mainly their level of participation, knowledge and skill improvements and gains in earnings. However, since many of the policy motivations for the LTC RAP have to do with improving the performance of and quality of care of long-term care providers, this section explores both possible studies of apprentices and their employers. The overarching approach for most designs is to compare the apprentice and employer performance to what it would have been in the absence of the LTC RAP. There are many possible ways to conduct that comparison. **Exhibit 1** provides a broad overview of the range of methods, types of data collection, and their relationship to costs and ability to generalize the findings to the total population of LTC RAPs. In some cases, the studies assess associations between the independent and dependent variables, while other studies make a stronger claim that the LTC RAP caused changes in the dependent variables.

EXHIBIT 1. Overview of Methods, Data Collection, and Potential Feasibility			
Analysis Methods	Qualitative Analysis	Quantitative Analysis	
		Descriptive analysis (single point in time AND no comparison group)	Multivariate analysis (two points in time OR with a comparison group OR both)
Data Collection	Focus groups of workers Case studies of employers In-depth ethnographic studies and implementation evaluation	Survey of LTC RAP workers or employers across occupation/ organization types LTC RAP administrative data Linking Medicare/ Medicaid claims/ OSCAR data	Survey only Survey and existing secondary data (National Nursing Assistant Survey; National Home and Hospice Care Survey)
Feasibility	Lower cost		Higher cost
	Lower generalizability		Higher generalizability

5.1. Outcome Variables

Exhibit 2 is a list of potential outcome variables for workers and employers for an evaluation of LTC RAPs. Some of these outcomes, such as wages and turnover, can be obtained from employers or from administrative records without having to ask the apprentices and the comparison group. Other outcomes, such as intent to leave, job

satisfaction, and relationship with supervisor can be obtained only by gathering directly through focus groups or surveys with the workers themselves. Finally, some outcomes, such as skill proficiency, would be exceptionally very hard to obtain because of the lack of agreed upon measures and the difficulty and expense of data collection. Data on skill proficiency would need to be obtained either through direct observation of individual workers, an opinion ranking by supervisors, or by the completion of some test, and could be challenged as biased or not measuring true skill levels.

EXHIBIT 2. Potential Outcome Variables for Workers and Employers	
Workers	<ul style="list-style-type: none"> • Earnings and fringe benefits • Job tenure, turnover, intent to leave job, intent to leave field • Job satisfaction • Satisfaction with employer • Relationship with supervisor, clients, and other staff • Advancement to more advanced jobs (e.g., are apprentices more likely to say that they will obtain additional training) • Participation in means-tested public programs (e.g., food stamps, Temporary Assistance for Needy Families [TANF], Medicaid) • Satisfaction with LTC RAP • Skill proficiency
Employers	<ul style="list-style-type: none"> • Offering of higher wages and more fringe benefits • Employer evaluation of apprentice skill development • Employer satisfaction with LTC RAP • Job tenure, turnover • Wages and fringe benefits provided • Provision of career ladder • Quality of care/Quality of Life • Net costs

Given their policy importance, quality of care outcomes at the level of the firm deserve special mention. Quality of care data in a standardized format is readily available for nursing homes and home health agencies, but not for other long-term care providers (Wiener, Freiman, and Brown, 2007). The Centers for Medicare and Medicaid Services (CMS) routinely posts detailed quality of care information for these two providers on their Nursing Home Compare and Home Health Compare web sites. Most of these measures are derived from resident and patient level data that is routinely and periodically collected on functional status and medical condition. In addition, for nursing homes, CMS regularly calculates summary measures, the Five Star Rating System, that includes information from the resident and patient assessments, staffing levels, and the health inspections. Data are not available on services for people with intellectual disabilities or for residential care facilities, such as assisted living facilities. Developing quality measures for these providers would be a major task and beyond the scope of a LTC RAP evaluation. In addition, even with the largest occupation, CNAs, there are currently only 56 employer sponsors, a sample size too small to detect differences in quality of care across nursing homes. Moreover, LTC RAPs are likely to have significant facility/agency-wide impacts only where apprentices account for a significant proportion of workers; it is probably unrealistic to expect a few apprentices in an organization to affect the overall quality of care.

5.2. Evaluation Designs to Determine Effects on Apprentices

Exhibit 3 presents a range of illustrative evaluation designs for the evaluation of the LTC RAP's effects on apprentices, ranked from strongest to weakest according to scientific rigor. The strength of the design is an assessment of how well the findings can be defended as measuring the true effect of the intervention as opposed to resulting from a methodological weakness in the design. The strongest designs have the highest costs; conversely, the lowest cost designs have major limitations. Although this exhibit ranks these potential evaluation designs in terms of their scientific rigor, it does not necessarily rank them in terms of their feasibility and desirability for the evaluation of the LTC RAP.

EXHIBIT 3. Possible Research Designs for Evaluation of Effects of LTC RAP on Apprentices, Ranked from Strongest to Weakest in Scientific Rigor		
Design	Pros	Cons
Randomly assign eligible applicants for long-term care positions to a LTC RAP or to a standard long-term care training program (randomized controlled trial) with data collection at assignment, when apprenticeship ends, and one year afterwards. Apprentices who do not complete the program would be followed	<ul style="list-style-type: none"> • Strongest possible design, with recognized ability to attribute effects to the intervention • Able to address a wide range of outcome variables 	<ul style="list-style-type: none"> • LTC RAPs and control group employers not likely to accept a randomized design because they lose control of an important component of their business • LTC RAP is already an ongoing program, not a demonstration • LTC RAP has too few employers for randomization • Requires at least two rounds of expensive data collection
Compare the changes over time in outcomes of apprentices entering the LTC RAP with entrants to standard long-term care jobs in other facilities/agencies (quasi-experimental design with a comparison group). Apprentices who do not complete the program would be followed	<ul style="list-style-type: none"> • Relatively strong design, commonly used in social science evaluations • Comparison groups could be made more similar through matching, propensity scoring, or multivariate analyses • Able to address a wide range of outcome variables 	<ul style="list-style-type: none"> • Results may be from unmeasured differences in workers or providers and not the apprenticeship program • No compelling reason for comparison group employers and workers to participate, reducing response rate • Requires at least two rounds of expensive data collection
Rigorous non-experimental methods, including natural experiments	<ul style="list-style-type: none"> • Offers methods for obtaining rigorous impact estimates without requiring employers to use random assignment • Yields estimates based on actual program operations not based on a change in approach necessitated by the evaluation • May go together with the administrative record option listed below 	<ul style="list-style-type: none"> • Natural experiments rely on events not under the control of the evaluator • No known natural experiments are currently available for a LTC RAP experiment. • Likely to be limited to a few sites, rather than a national evaluation

EXHIBIT 3 (continued)		
Design	Pros	Cons
Use administrative records, such as Unemployment Insurance data, to examine how earnings and retention of long-term care workers who participated in LTC RAP compare with long-term care workers in other agencies/ facilities who have not participated in LTC RAP, controlling for work history, previous earnings, and other available variables. (Quasi-experimental design with comparison group). All persons who ever participated in the LTC RAP program would be included, regardless of whether they completed the program.	<ul style="list-style-type: none"> • Low data collection costs if researchers can gain access to Social Security numbers and Employer Identification Numbers, which should be possible with proper protections • No new reporting burden on employers or workers • Offers direct measures of turnover and earnings • Earnings and turnover information will be relevant to any cost-benefit assessment 	<ul style="list-style-type: none"> • Limited matching or control variables available in administrative datasets • Cannot control for unmeasured differences between the two groups • From case studies, wage rates are unlikely to increase much. Turnover for apprentice population ambiguous measure because it may signify that workers left for more formal education required to advance in field • Outcome variables limited to what is in administrative dataset • No data on views of workers or employers
Compare apprentices when they begin their training after 1-2 years (pre/post design). Only new apprentices would be included, but they would be followed regardless of whether they completed the program	<ul style="list-style-type: none"> • Does not require recruitment of comparison group • Lower cost than gathering information for separate treatment and comparison group • Able to address a wide range of outcome variables 	<ul style="list-style-type: none"> • Changes cannot be definitively attributed to LTC RAP because design does not control for secular trends, such as recessions, inflation, changes in demand for services, and increasing experience of workers • Requires long data collection period as new apprentices enroll in LTC RAP on a flow basis, driving up costs
Compare apprentices at a point in time with workers who did not participate in LTC RAP in other agencies/ facilities. Persons who had ever been in the apprentice program and were still working for the same employer would be included	<ul style="list-style-type: none"> • Relatively low-cost because only one round of data collection 	<ul style="list-style-type: none"> • Relatively weak design because there is no comparison of change in outcomes over time • Since cross-sectional analysis, cannot say that differences were “caused” by LTC RAP; can only say that there is an “association” between variables • Cannot fully control for potential selection bias
Collect single wave of data from apprentices with focus on comparing subgroups, such as Whites versus ethnic/ racial minorities. Persons who had ever been in the apprentice program and were still working for the same employer would be included	<ul style="list-style-type: none"> • Lower cost because only single wave of data collection • Obtain information on views of apprentices on LTC RAP, which may be useful for program improvement 	<ul style="list-style-type: none"> • Cannot answer question of program effectiveness because no comparison to people who did not participate in program • Does not help build “business case” for apprenticeship programs

EXHIBIT 3 (continued)		
Design	Pros	Cons
Conduct focus groups of workers who are apprentices and of workers who are not apprentices. Persons who had ever been in the apprentice program and were still working for the same employer would be included	<ul style="list-style-type: none"> • Low-cost • Provides detailed views of workers • Can provide detailed recommendations for improving LTC RAP • Allows for some comparison with workers not in apprenticeship programs 	<ul style="list-style-type: none"> • Qualitative data cannot be used to determine effectiveness of intervention • Representativeness of views expressed cannot be directly assessed • Comments provided cannot be easily summarized or quantified
Conduct focus groups only with apprentices. Persons who had ever been in the apprentice program and were still working for the same employer would be included	<ul style="list-style-type: none"> • Lowest cost option • Provides information on views of apprentices • Can provide detailed recommendations for improving LTC RAP 	<ul style="list-style-type: none"> • Qualitative data cannot be used to determine effectiveness of intervention • Representativeness of views expressed cannot be directly assessed • Views by apprentices provided cannot be easily summarized or quantified • Comparisons cannot be made to workers who did not participate in LTC RAP

5.3. Evaluation Designs to Determine Effects on Employers

Exhibit 4 presents a range of research designs for the evaluation of the program's effects on employers and pros and cons for each design. These designs are not necessarily mutually exclusive and could be combined in various ways. The relatively small number (119 across four occupational categories) of programs limits the evaluation design options for employers, but new LTC RAPs may expand options. In addition to gathering information on the effect of the apprenticeship programs on employers, data on employers would also provide control variables for the analyses of the effect of LTC RAP on direct care workers.

EXHIBIT 4. Research Designs for Evaluation of Effects of LTC RAP on Employers, Ranked from Strongest to Weakest in Scientific Rigor		
Design	Pros	Cons
Randomly assign long-term care employers into two groups: a treatment group that will heavily use LTC RAP and a control group that will not use LTC RAP	<ul style="list-style-type: none"> • Strongest possible design, with recognized ability to attribute effects to the intervention 	<ul style="list-style-type: none"> • Obtaining a large enough sample of employers to do quantitative analysis will be difficult and expensive • Providers recruited because they are interested in improving their training programs may not be satisfied with being in control group and may adopt other training programs • Some providers in the intervention group may not implement LTC RAPs • Requires a long time period for employers to adopt the program and master its use

EXHIBIT 4 (continued)		
Design	Pros	Cons
Compare outcomes for multi-site employers who use LTC RAP in some sites but not in other sites; also compare changes in outcomes by site	<ul style="list-style-type: none"> • Holds constant many employer-specific factors not linked to the type of training • Might attract employer participation and interest • Captures worker and employer impacts, including potential organizational effects • Offers a direct way of estimating costs and benefits to the employer 	<ul style="list-style-type: none"> • Number of LTC RAP employers which have multiple sites is small • Might involve selection bias because sites that employers choosing to implement LTC RAP may be systematically different from sites not using long-term care • Outcomes might depend non-LTC RAP site-specific factors • If number of apprentices per employer is small, program is not likely to have an impact on organizational performance
Compare outcomes for LTC RAP employers with matched employers not offering LTC RAPs	<ul style="list-style-type: none"> • Relatively strong design, commonly used in evaluations • Captures worker and employer impacts, including potential organizational effects • Offers a direct way of estimating costs and benefits to the employer 	<ul style="list-style-type: none"> • Results may be result of selection bias, if employers who are more (or less) effective in other ways disproportionately adopt LTC RAP • Cross-sectional design limits ability to interpret differences as due to the intervention • Comparison group may include facilities with some other training initiative • If number of apprentices per employer is small, program is not likely to have an impact on organizational performance
Cost-benefit analysis to ascertain costs for local LTC RAP design and implementation and benefits realized as improvements in worker productivity and quality	<ul style="list-style-type: none"> • Low-cost • Commonly used in social science evaluations • Requires only a modest number of employers 	<ul style="list-style-type: none"> • Does not employ statistical controls found in regression approaches listed above for non-LTC RAP effects and selection bias • Less generalizable than regression-based approaches -- no comparison group • Requires significant cooperation of employers to provide data
Conduct focus groups of employers who operate LTC RAPs and of employers who do not offer LTC RAPs. Participants will be primarily administrators attending national conferences	<ul style="list-style-type: none"> • Low-cost • Provides detailed views of employers • Can provide detailed recommendations for improving LTC RAP • Allows for some comparison with employers not offering LTC RAPs 	<ul style="list-style-type: none"> • Qualitative data cannot be used to yield quantitative impact estimates of the effectiveness of intervention • Representativeness of views expressed cannot be directly assessed • Comments provided cannot be easily quantified and counted

EXHIBIT 4 (continued)		
Design	Pros	Cons
Conduct focus groups only with employers operating LTC RAPs	<ul style="list-style-type: none"> • Low-cost option • Provides information on views of employers • Can provide detailed recommendations for improving the LTC RAP • Focus groups could be organized at national provider conventions 	<ul style="list-style-type: none"> • Qualitative data cannot be used to determine effectiveness of intervention • Representativeness of views expressed cannot be directly assessed • Comments provided cannot be easily quantified and counted • Comparisons cannot be made to employers who do not participate in LTC RAP
Case studies of LTC RAP employers compared to employers without LTC RAPs	<ul style="list-style-type: none"> • Low-cost • Can obtain “rich” description of programs and employer views • Offers comparison to organizations without programs 	<ul style="list-style-type: none"> • Qualitative data cannot be used to determine effectiveness of LTC RAP • Results may be result of selection bias, if employers who are more (or less) effective in other ways disproportionately adopt LTC RAP • Case studies recently have been conducted of major LTC RAPs. Relatively little to be gained by more case studies at this time
Case studies of LTC RAP employers with no comparison group	<ul style="list-style-type: none"> • Lowest cost • Can obtain rich description of programs and employer views 	<ul style="list-style-type: none"> • Qualitative data cannot be used to determine effectiveness of LTC RAP • No comparison group • Case studies recently have been conducted of major LTC RAPs. Relatively little to be gained by more case studies at this time

5.4. Data Sources

To conduct an evaluation of this type, several qualitative and quantitative methods and sources of data could be used. These include:

- Surveys that would gather systematic, quantitative information on large numbers of people or programs through mail questionnaires or telephone, web or in-person interviews. At the extremes, mail or web surveys are the lowest cost, but also have the lowest and often biased response rates; in-person surveys are the most expensive approach. Web and mail surveys (with follow-up) might work effectively for surveys of employers. Telephone surveys are typically in between in terms of cost, but the increase in the use of cell phones and the decline in the use of landlines often make these surveys problematic for younger and lower income populations. All of these surveys require relatively large numbers of names and contact information, such as mailing address, telephone number, or e-mail address, which is often difficult to obtain, especially for comparison groups. Experience with the National Nursing Assistant Survey and with the

National Home Health Aide Survey suggests that approximately 76% of nursing homes (71% of home health agencies) would respond and within nursing homes 70% of apprentices and other workers (79% of home health apprentices) would respond for an overall two stage response rate of approximately 53% (Squillace, Remsberg, and Bercovitz, 2006; National Center for Health Statistics (NCHS), undated).

- Administrative datasets, such as Unemployment Insurance quarterly earnings record and CMS quality of care data, would provide useful information on factors such as wages, job tenure, job history and facility/agency performance without having to survey respondents. Administrative datasets are collected for purposes other than research and have limited variables on the characteristics of workers and employers. DOL has the Social Security numbers of people who have participated in LTC RAPs, which would allow the identification of apprentices in these databases. Privacy concerns may limit what information can be released since the number of apprentices and employers is relatively small and possibly could be identified in the data. At the employer level, as noted above, CMS has a large amount of quality of care data about individual nursing homes and home health agencies that is publicly available in downloadable datasets. Similar information is not available for programs for people with developmental disabilities or for residential care facilities.
- Focus groups, which are structured conversations with small groups of respondents (e.g., apprentices or employers) about issues of interest for the evaluation. This approach provides detailed information on the perspectives of a relatively small number of people. The information does not provide quantitative data and cannot be used to determine the effectiveness of LTC RAPs.
- Case studies, which would include structured discussions with multiple stakeholders in a LTC RAP. For a case study of a LTC RAP, the evaluators would interview the agency or facility administrator, LTC RAP director or liaison, mentors, instructors, apprentices, supervisors, and state officials. If implemented, these case studies would build on the site visits conducted for this contract (Kuehn et al., 2011). However, the RTI International/Urban Institute team conducted detailed case studies in 2011 with almost all of the larger programs, so additional case studies may not substantially add to the information already available.

5.5. Control or Comparison Groups

The essence of evaluation research is to answer the question: How do the outcomes of the participants in the intervention compare with the outcomes they would have experienced had they not participated? For this evaluation, the focus is on how the LTC RAP training affects workers and employers compared to the standard training that direct care workers normally receive.

A key issue, then, is how to identify a comparison group that is similar to apprentices in the LTC RAP. In the case of random assignment, the control group is very likely to yield reliable estimates of what would have happened in the absence of the intervention (also known as the “counterfactual”) because the assignments to treatment and control groups are random. However, for other types of evaluations, a potential threat to the validity of the findings is that the comparison group and the LTC RAP group differ in terms of characteristics that affect the outcomes of interest but are not related to the operation of the intervention. This could well be the case if apprentices are more motivated and ambitious than people who are not in the LTC RAP. Similarly, employers that voluntarily participate in the LTC RAP are likely to have an organizational culture that places a higher priority on the importance of training direct care workers than employers that do not participate in the program.

An additional complication in choosing comparison groups is that the effect of the LTC RAP is likely to vary over time; that is, the effect of the program on an apprentice is likely to be greater towards the end of the training program than it is at the very beginning. Thus, ideally, the comparison group would start with people at the beginning of their employment at the provider and follow them over time. However, collecting data on a flow basis is more difficult and expensive than data collected from people all starting at the same time because data need to be collected whenever a new person starts work, which could be over a long period of time.

Some possible comparison groups are:

- *Randomized treatment and control groups.* The classic solution to the problem of selection bias is randomization between the treatment and control groups. However, randomization is more difficult to implement in ongoing programs because employers may be reluctant to randomly assign workers to an apprenticeship if they do not think they are capable of benefiting from it. Similarly, they may be unwilling not to assign workers to the treatment group if they think the workers could benefit the organization by receiving the apprenticeship training. In addition, employers may not be willing to be part of a control group if their motivation in participating in the study is to improve their training programs.
- *Matched sample of employers and matched sample of workers working for these employers.* While matching employers based on simple characteristics, such as number of nursing home beds, would be simple; gathering information about other variables, such as organizational culture, would be difficult and time consuming. Employers who do not have an apprenticeship program would have limited incentives to participate in the survey and to provide confidential contact information on their employees. Two levels of sampling -- facility/agency and workers -- could result in relatively low final response rates.
- *Apprentices with matched workers within the same employer.* It would be easier to identify the sample and compare people without having to control for the

employer. However, if LTC RAP participation causes changes for the organization as a whole (or if employers choose to participate because of certain employer characteristics), then differences between apprentices and non apprentices may be minimal. Moreover, findings from the site visits suggest that employers consciously use apprentices as peer-mentors for persons not in the apprenticeship program, thus “contaminating” workers who are not formally in the apprenticeship program. This approach would also leave the evaluation without an employer-level comparison.

5.6. Evaluation of Broad Options

In reviewing the options presented in this section, it is necessary to weigh the importance of the questions to be asked, the feasibility of the approach, the scientific rigor of the evaluation design, and the cost of implementing the design. Ultimately, the decision about whether to implement an evaluation and which one is a question of value for money -- does the value provided by the evaluation merit the expenditure of the funds?

In terms of feasibility, scientific rigor, and cost, the RTI International/Urban Institute team believes that some of the options presented above seem especially weak at this time. First, because case studies have recently been conducted of five of the largest LTC RAPs as part of this project, there is little to be gained in conducting additional case studies at this time. Most of the remaining programs are relatively small. If the evaluation is not conducted until several years from now, then case studies may be worthwhile. Second, while randomized controlled trials are the gold standard of research, the LTC RAP is already an ongoing intervention. A randomized controlled trial would be difficult to implement on a scale large enough to yield statistically significant results and would be expensive. While an attractive option in many ways, it does not seem to meet the mandate of evaluating an ongoing program. It could only be considered seriously in the context of a major demonstration program and the likely need for significant government or foundation funding. Third, pre-post designs without comparison groups seem particularly problematic for job training programs where merely continuing to work at the job would provide individuals with increased experience and expertise in caregiving even without any formal additional training. Moreover, this approach cannot control for events external to the training, such as inflation, recessions, and changes in management, which may affect the performance of direct care workers. Fourth, although a strong research design, new data collection that involves following a cohort of apprentices and a comparison group from the beginning of their training through the end of the apprenticeship and perhaps some time later would be expensive because of the long time it would take to gather data on a sufficient number of apprentices and comparison workers and the long training period. Data collection on a flow basis is expensive.

6. DETAILED DESCRIPTION OF FOUR APPROACHES TO EVALUATING THE LTC RAP

In this section, we describe a four component approach to evaluating the LTC RAP. The four components are: (1) use of the Longitudinal Employer Household Dynamics (LEHD) administrative dataset to compare apprentices with non-apprentices and LTC RAP employers and non-employers; (2) a cross-sectional telephone survey of apprentices and non-apprentices; (3) focus groups with apprentices and employers without a comparison group; and (4) a cost-benefit analysis of the LTC RAP from the employer's perspective. With the exception of the cost-benefit analysis, which depends in part on the analyses of the administrative dataset and the telephone survey, each component is separate, but complementary, and could be funded without the others.

6.1. LEHD Quasi-Experimental Evaluation Option

This option for evaluating the LTC RAP relies on using RAPIDS administrative data on LTC RAP apprentices combined with employer and employee information from the LEHD database, a Census Bureau data file that includes state-level Unemployment Insurance administrative information on employment and earnings. This quasi-experimental strategy identifies a comparable comparison group for LTC RAP apprentices by matching characteristics of apprentices with other long-term care workers in the database who have not participated in the apprenticeship program. Similar matching and comparison is also performed for employer sponsors.

Research Questions

This evaluation design helps answer the following research questions:

- What is the impact of registered apprenticeship on apprentices?
 - How does participation in a LTC RAP affect job tenure within an employer, compared to tenure and employment stability in the absence of the apprenticeship program?
 - How does participation in a LTC RAP affect employment stability within the long-term care industry, compared to employment stability within the long-term care industry in the absence of the apprenticeship program?
 - How does participation in a LTC RAP affect earnings growth compared to earnings growth in the absence of the apprenticeship program?
- What is the impact of registered apprenticeship on LTC RAP sponsors?
 - How does offering a LTC RAP affect overall direct care worker turnover, compared to worker turnover in the absence of the apprenticeship program?

- How does offering a LTC RAP affect the loss of employees to competing long-term care providers, compared to the loss of employees in the absence of the apprenticeship program?
- How does offering a LTC RAP affect employment and revenue growth, compared to employment and revenue growth in the absence of the apprenticeship program?

Overview of Evaluation Design

This quasi-experimental design would use firm and worker level LEHD data on apprenticeship sponsors to construct treatment groups of workers who have participated in apprenticeship programs and comparison groups of workers for non-sponsor long-term care providers. The LEHD is a research program directed by the Census Bureau that combines state administrative data on workers and firms from state Unemployment Insurance programs to generate a national matched quarterly employer-employee database. Because this data is required by law, it is unusually complete and accurate on the limited variables it includes. Unlike administrative data available from individual states, the LEHD data can identify workers who move across state lines and firms who have workers in multiple states. Since the core of the LEHD is a matched employer file and worker file, it captures labor market dynamics from both the firm and the workers' perspective. Another feature of the LEHD is that a variety of other Census surveys and Internal Revenue Service data are matched to employers and employees using Social Security numbers and Employer Identification Numbers. These identifiers are not directly available to researchers, but researchers can submit lists of Social Security numbers and Employer Identification Numbers to the Census Bureau to obtain an extract of the LEHD data for analysis.

To answer research questions related to the impact of the LTC RAP on apprentices, the following treatment and comparison groups will be constructed from the LEHD data:

- All workers who have participated in LTC RAPs and all low-wage workers in matched long-term care firms not administering a LTC RAP.
- All workers who have participated in LTC RAPs and all non-apprentice low-wage workers in LTC RAP sponsoring long-term care firms.
- All low-wage workers employed by LTC RAP sponsors and all low-wage workers in matched long-term care firms.

One limitation of the LEHD data in constructing these comparison groups is that low-earning direct care workers employed by long-term care firms cannot be distinguished from other low-earning workers employed in long-term care firms, such as housekeeping staff and dietary workers. However, if a known apprentice is matched to housekeeping/dietary staff in a sample, it will be because their earnings histories are extremely close. This suggests a certain degree of substitutability between occupations.

Of course there may be some unobserved motivations and values for these different workers, but what matters is the question “*if these apprentices had not received an apprenticeship would their wage trajectories be comparable to the earnings trajectories of someone in the same industry with the same earnings history.*” This limitation is an issue primarily for nursing homes, residential care facilities and group homes for people with intellectual disabilities because they provide room and board and employ substantial numbers of housekeeping and dietary staff members; it is less of a problem for home health agencies because they hire relatively fewer non-direct care staff.

To answer research questions related to the impact of the LTC RAP on employers, the treatment group would be all LTC RAP sponsoring firms and the comparison group would be non-LTC RAP sponsoring long-term care firms matched to the treatment groups on variables, such as geographic location, establishment revenue, and number of low-wage employees using propensity score matching techniques. Propensity score matching uses a statistical model to predict the probability of being in the treatment (apprenticeship) group using a series of observable characteristics. This predicted probability of being in the treatment group is then used to construct weights for the comparison group, which make the weighted comparison group more comparable to the treatment group on observable characteristics, therefore somewhat mimicking random assignment. The impact of registered apprenticeship on employment, earnings, turnover, job growth, and worker separation outcomes can be estimated comparing the difference-in-means between the treatment and comparison groups after matching or, ideally, as a difference-in-differences multivariate model, which compares the differences in the outcome over time for apprentices or firms offering LTC RAP and firms not administering a LTC RAP (a strategy that Mueser, Troske, and Gorislavsky (2007) find to be superior in program evaluations of job training programs using propensity score matching of administrative data).

An important advantage of this evaluation design approach is that it can use all persons who have ever been apprentices in the LTC RAP, including those who no longer work for the long-term care provider that administered the apprenticeship or even those no longer working in long-term care. Since the identification of a comparison group is completed after the identification of LTC RAP apprentices, appropriate comparison groups can be constructed for all LTC RAPs, regardless of their implementation date. Indeed, the use of RAPIDS data over a several year period and over several sponsors helps ensure that outcomes are not affected by cyclical factors, and circumstances in specific local markets. This cannot be guaranteed by an experimental design at a specific site. **Exhibit 5** summarizes the proposed LEHD impact analyses.

EXHIBIT 5. Summary of LEHD LTC RAP Impact Analyses			
Research Question	Treatment Group	Comparison Group	Outcome Variable
What is the impact of registered apprenticeship on long-term care workers?	All LTC RAP apprentices, regardless of whether they completed the program or whether they currently still work for the same employer or in a long-term care setting	Low-wage workers in non-LTC RAP long-term care firms, matched with LTC RAP apprentices	<ul style="list-style-type: none"> Continued employment at sponsor Earnings Tenure in the long-term care industry
What is the impact of registered apprenticeship on long-term care workers?	All LTC RAP apprentices, regardless of whether they completed the program or whether they currently still work for the same employer or in a long-term care setting	Low-wage, non-apprentice workers in LTC RAP sponsoring long-term care firms, matched with LTC RAP apprentices	<ul style="list-style-type: none"> Continued employment at sponsor Earnings Tenure in the long-term care industry
What is the impact of registered apprenticeship on long-term care workers?	All low-wage workers employed by LTC RAP sponsors	Low-wage workers in non-LTC RAP long-term care firms, matched with all low-wage workers employed by LTC RAP sponsors	<ul style="list-style-type: none"> Continued employment at sponsor Earnings Tenure in the long-term care industry
What is the impact of registered apprenticeship on long-term care employers?	All LTC RAP sponsors	All non-registered apprenticeship program sponsoring long-term care firms, matched with LTC RAP sponsors	<ul style="list-style-type: none"> Turnover Employment growth Loss of workers to other long-term care providers Revenue growth

Treatment of Long-Term Care Occupations

Occupations are identified in the RAPIDS database with occupational codes. In contrast, the LEHD data do not identify worker occupations, but they do identify the industry of the employer. This would typically be an obstacle to identifying an appropriate comparison group, but in the case of long-term care there is a close correspondence between occupation and industry groups. The predominant LTC RAP occupations are CNAs, HHAs, HSSs and DSSs. These occupations align with the industry sectors presented in ***Exhibit 6***.

EXHIBIT 6. Occupation-Industry Crosswalk			
Occupation	RAPIDS Occupation Code	Industry	Four Digit NAICS Industry Code
Certified Nursing Assistants (CNA)	824, 824C, 824CB, 824A, 824R, 824D, 824G, 824M	Nursing Care Facilities	6231
Home Health Aides (HHA)	1086, 1086CB, 1086A, 1086B, 1086D, 1086E	Home Health Care Services	6216
Health Support Specialists (HSS)	1086AA	Assisted Living Facilities, and Other Residential Care	6233, 6232, 6239
Direct Support Specialists (DSS)	1040, 1040CB	Services for Elderly & Persons with Disabilities	6241
NOTES: NAICS Code 6231: Nursing Care Facilities; 6216: Home Health Care Services; 6233: Continuing Care Retirement Communities and Homes for the Elderly; 6232: Residential Mental Health and Substance Abuse Facilities; 6239: Other Residential Care Facilities; 6241: Children and Youth Services, Services for the Elderly and Persons with Disabilities, and Other Individual and Family Services.			

Treatment cases of a specific occupation in the RAPIDS data can be matched to corresponding comparison cases employed by a firm in the corresponding industry in the LEHD data in two different ways. First, treatment cases could be segregated by industrial sector, and matched exclusively to comparison cases in the same sector, so that the propensity score matching is done separately by industry group (which is expected to correspond closely to the occupation group of interest). However, some long-term care firms provide several different services, but are required by state administrative data systems to report only one industry (typically their predominant activity). For example, a nursing home facility operated by a hospital might be classified as a “hospital” rather than a “nursing home” for its industry codes, because the hospital facility is the firm’s predominant activity. To account for this, a second option is to consider all treatment cases (i.e., not segregate them by industry), match them to all comparison cases, and use industry groups as a matching variable only, rather than as a way of identifying industrial sub-samples.

Sample Design

Potential selection bias in this quasi-experimental evaluation of the LTC RAP could occur in at least two ways: the non-random selection of different providers into the LTC RAP initiative or the non-random selection of different employees into the LTC RAP within the provider. One possibility is that higher quality, more financially secure long-term care providers are more likely to start a LTC RAP. On the other hand, those long-term care providers already doing well and satisfied with their training programs may be least likely to use the LTC RAP. From this perspective, bias could run in either direction, especially if the comparison is based on levels and not on changes for each group.

Also, site visits indicate that higher quality employees are generally chosen to participate in the LTC RAPs as apprentices. If so, estimates based on simple comparisons of participants and non-participants might overstate the impact of LTC RAP. To address this problem, an evaluation can use propensity score matching on pre-apprenticeship enrollment earnings (which should help to capture unobservable human capital that contributes to on-the-job productivity), age, gender, job tenure, firm size, and industry to identify a comparison sample. Pre-program earnings and employment records may be a good matching indicator, one likely to capture individual differences in unmeasured pre-program characteristics related to performance in the job market. While variation in *wage rates* is modest for long-term care workers, there is considerable variation in hours worked and in job tenure. Since the LEHD only provides information on quarterly earnings (i.e., hourly wages multiplied by hours worked), there should be more variation in earnings than in wages. This method is widely used by evaluators of other programs targeted at low-wage workers (Mueser, Troske, and Gorislavsky, 2007).

Access to the universe of long-term care providers and LTC RAP sponsors in state Unemployment Insurance administrative data systems through the LEHD allows for the construction of a variety of comparison and treatment groups. Testing between multiple

treatment and comparison groups enables evaluators to address biases that may be present in some specifications, but not others. For example, potential selection bias may produce a regression estimate either above or below the true effect. Using two different comparison groups for each outcome will allow generation of these upper and lower bound effects so that the true effect can be bounded between the two estimates. This strategy provides some assurance of the range of the treatment effect.

Multiple treatment/comparison group pairs are amenable to the propensity score matching approach, and each pair has advantages and disadvantages associated with it. These are summarized in **Exhibit 7**. If propensity score matching on observable characteristics successfully accounts for unobservable characteristics of apprentices and LTC RAP sponsors because they are correlated with the unobservable characteristics, then selection biases can be minimized.

EXHIBIT 7. Advantages and Disadvantages of LEHD Impact Analyses			
Treatment Group	Comparison Group (to be matched to the Treatment Group)	Advantages	Disadvantages
All RAPIDs apprentices	All low-wage workers in non-LTC RAP long-term care firms	<ul style="list-style-type: none"> • Attempts to control for selection bias associated with selection of <i>apprentices</i> on unobservable characteristics within the firm • Large sample size • Able to use all apprentices who ever were in a LTC RAP • Able to follow apprentices to other employers within and outside long-term care • Addresses key issues of job tenure and income growth 	<ul style="list-style-type: none"> • May not capture benefits that spillover to non-apprentice employees of a LTC RAP sponsor • May suffer from selection bias associated with selection of <i>sponsors</i> into the LTC RAP • Limited variables on which to match employees and employers, which may result in selection bias
All RAPIDs apprentices	All low-wage workers in LTC RAP sponsoring firms	<ul style="list-style-type: none"> • Does not suffer from selection bias associated with selection of <i>sponsors</i> into the registered apprenticeship program • Large sample size • Able to follow apprentices to other employers within and outside long-term care • Addresses key issues of job tenure and income growth 	<ul style="list-style-type: none"> • Suffers from contamination bias to the extent that benefits of apprenticeship spillover to non-apprentice employees of a LTC registered apprenticeship program sponsor • May suffer from selection bias associated with selection of <i>apprentices</i> on unobservable characteristics within the firm

EXHIBIT 7 (continued)			
Treatment Group	Comparison Group (to be matched to the Treatment Group)	Advantages	Disadvantages
All low-wage workers employed by LTC RAP sponsors	All low-wage workers in long-term care firms, regardless of whether they completed the program or whether they currently still work for the same employer or in a long-term care setting	<ul style="list-style-type: none"> • Captures benefits that spillover to non-apprentice employees of a registered apprenticeship program sponsor • Does not suffer from selection bias associated with selection of <i>apprentices</i> on unobservable characteristics within the firm • Large sample size • Able to follow apprentices to other employers within and outside long-term care • Addresses key issues of job tenure and income growth 	<ul style="list-style-type: none"> • May suffer from selection bias associated with selection of <i>sponsors</i> into the registered apprenticeship program • Dilution of treatment effect by inclusion of other provider employees who do not provide direct care
All LTC RAP sponsors	All non-LTC RAP sponsoring long-term care firms	<ul style="list-style-type: none"> • Capture benefits that spillover to non-apprentice employees of a registered apprenticeship program sponsor 	<ul style="list-style-type: none"> • May suffer from selection bias associated with selection of <i>sponsors</i> into the registered apprenticeship program • Lack of statistical power

If propensity score matching on observable characteristics successfully accounts for unobservable characteristics of apprentices and LTC RAP sponsors because they are correlated with the unobservable characteristics, then selection biases can be minimized.

1. Treatment Groups

The treatment groups used in the evaluation, presented in **Exhibit 8**, will be drawn from employers and apprentices in the RAPIDS data that are identifiable in the LEHD data. The LEHD covers all workers covered by the state Unemployment Insurance program (which should be the entire RAPIDS universe). A total of over 4,300 unique LTC RAP participants are included in the RAPIDS system between January 2005 and May 2011, representing 119 programs. This treatment group may expand if more LTC RAPs are implemented between May 2011 and any evaluation. If sample sizes were adequate, separate analyses could be conducted of workers who completed the apprenticeship program, people currently in the program, and people who dropped out prior to completing the program. It is not uncommon for social programs to have high dropout rates and be effective for those who complete the program.

EXHIBIT 8. Treatment and Comparison Group for LEHD Analyses	
Treatment Group	Comparison Group
All RAPIDs apprentices (N~3,750)	All low-wage workers in long-term care firms (N~2,000,000)
All RAPIDs apprentices (N~3,750)	All low-wage workers in LTC RAP sponsoring firms (N~5,000)
All low-wage workers employed by registered apprenticeship program sponsors (N~5,000)	Low-wage workers in long-term care firms (N~2,000,000)
All registered apprenticeship program sponsors (N~119)	All non-registered apprenticeship program sponsoring long-term care firms (N~100,000)

2. Comparison Groups

If all LTC RAPs were implemented simultaneously, a single comparison group could be chosen for all programs. However, this is not the case. In order to guarantee that pre-apprenticeship characteristics of the treatment group are matched to characteristics of the comparison group during the same time frame, propensity score matching must be conducted separately for each quarterly wave of LTC RAP registration. Thus, the apprentices that register with a LTC RAP at varying points in time will be matched to comparison cases from the LEHD data on the basis of quarterly earnings occurring before the LTC RAP registration.

The propensity score matching to determine the appropriate weights for the comparison group would follow Rubin (2001) and Mueser, Troske, and Gorislavsky (2007), and consist of an estimation of the predicted probability of being in the treatment group as a Logit function of eight quarters of earnings data and a set of additional variables, including industry/occupation group and geographic location for each quarter of the LEHD data available.

The propensity score matching approach is applied in two different ways to obtain the different comparison groups desired. In the first way, one applies propensity score weighting to construct a comparison group of roughly the same size as the treatment group (e.g., apprentices). In the second way, one applies weights to the population at large from whom the comparison group is drawn (e.g., all long-term care employers). Apprentices or employers/sponsors which have a close match get a high weight and apprentices or employer/sponsors which do not have a close match get a low weight.

Estimated Statistical Power

Evaluators develop statistical power calculations in order to determine the minimum sample size necessary to have confidence in the study's ability to detect a policy relevant impact. Interest in identifying the minimum sample size needed is common because of the cost of obtaining a larger sample, for example, by increasing the number of people required to complete surveys. For this particular evaluation, the evaluator would have access to a large sample at very modest cost. To be conservative, for the apprentice-level calculations, we used the approximate number of apprentices (3,750) in the RAPIDS data at the end of 2009 for the size of each of the treatment and control groups in the comparisons to be made. For the employer sponsor

calculation, we assumed 150 employer sponsors would have a LTC RAP by the time a potential evaluation was fielded; in other words, we assumed that additional programs would be added to the current 119 employer sponsors.

In estimating power calculations, the issue is: “How small a difference in each outcome measure (e.g., the “impact”) can be detected at $p < 0.05$ with approximately 80% power given the number of apprentices (or employer sponsors) in the analysis for the variation (expressed as the standard deviation) in the outcome measure?” Statistical power is the degree of confidence (probability) with which one can correctly reject the hypothesis that there is no impact. Conventionally, statisticians suggest that power of 80% is satisfactory.

We estimated statistical power for three different outcomes -- apprentice annual earnings and job tenure, and employer sponsor-level turnover. For the outcome measure of apprentice annual earnings, the analysis could detect a difference as small as \$300 in annual earnings using mean annual earnings of \$21,000 and a standard deviation of \$5,000, and assuming 3,750 employees each in the apprentice and comparison groups. For the outcome measure of apprentice job tenure, the analysis could detect a difference in job tenure as small as 0.7 months using mean tenure of 30 months and a standard deviation of 12 months, assuming 3,750 employees each in the apprentice and comparison groups. For the outcome measure of annual employer-sponsor turnover (where turnover is expressed in percentage points), the analysis could detect a difference as small as 5 percentage points using a mean turnover rate of 55 percentage points and a standard deviation of 25 percentage points, assuming 150 employers in the LTC RAP employer sponsor group and almost 99,850 employers in the comparison group. In each case, the evaluator would be able to detect even relatively low impacts of the LTC RAP.

Domains on Which Information Will Be Gathered

Information will be collected on treatment and comparison cases using data from the LEHD. An advantage of using this data is that it ensures that information is collected consistently across all programs, and between treatment and comparison groups. The primary domains on which information will be gathered are:

- **Quarterly earnings:** Earnings recorded in state unemployment insurance data systems are reported in the LEHD for each job held in a quarter.
- **Quarterly employment:** Cases will be considered employed in a quarter if they have positive earnings during that quarter. Quarterly employment information can also be used to construct a job tenure variable, which can be used for matching.
- **Industry:** The LEHD records four digit NAICS industry codes which will be mapped on to occupational codes in the RAPIDs data (see ***Exhibit 6***).

- **Employer:** Employer Identification Numbers are also provided in the LEHD data so that in addition to assessing the impact of the LTC RAP on employment and earnings in general, attachment to the RAP sponsoring firm, and firm turnover will also be determined.
- **Geographic location:** The geographic location of long-term care providers will be an important matching variable, which ensures that long-term care providers are compared to cases operating in comparable long-term care markets.
- **Demographic characteristics:** Age and gender are available on employees in the LEHD and can also be used to match. Education level and race/ethnicity are not available.
- **Firm revenue:** Gross revenues collected in economic censuses and linked to the LEHD can be used for matching to ensure that treatment cases are compared to comparison cases from similar sized firms.

In addition to this primary information, other firm-level data available in the LEHD may be used to improve the quality of the match between treatment and comparison groups. For example, information on firm age, if complete, could contribute to the quality of the match.

Data Collection Process

While no primary data collection will be necessary for this evaluation design, data will be obtained from the LEHD, which is maintained by the Census Bureau. Researchers must apply to the Census Bureau for use of the LEHD data for specific projects. This application process can be time consuming, and should be initiated very early in the evaluation process. Evaluators will not be able to identify data linked to specific Social Security numbers or employer identification numbers, but they will be able to submit these identifiers to the Census Bureau, so that cases can be extracted and assigned an identification number for the analysis.

Social Security numbers and Employer Identification Numbers for LTC RAP sponsors and participants will be drawn from the RAPIDs data for submission to the Census Bureau. After applying for use of the LEHD data and signing a data user's agreement, the Census Bureau will provide:

- All LEHD cases that have Social Security numbers that match the Social Security numbers of LTC RAP participants. These Social Security numbers will be replaced with personal identification keys and an indicator variable identifying the cases as apprentices.
- All LEHD cases that do not have Social Security numbers that match the Social Security numbers of registered apprenticeship program participants, but who have been employed by firms that have sponsored LTC RAPs. These firms will

be identified by an Employer Identification Number submitted to the Census Bureau. Social Security numbers for these cases will be replaced with personal identification keys and an indicator variable identifying these cases as non-apprentices.

- All LEHD cases who have been employed by firms that have not sponsored LTC RAPs but who have reported NAICS industry codes associated with the long-term care industry (Codes 6231, 6216, 6233, 6232, 6239, and 6241).

Time Frame to Collect and Analyze Data

We anticipate that the LEHD analysis option would take approximately 27 months to complete. The activities would include initial planning and data acquisition, including obtaining Social Security numbers for matching RAPIDS to LEHD data (15 months), data cleaning and analysis (6 months), and report development (6 months). According to DOL officials, the application process for obtaining personal data, such as Social Security numbers, takes about a year.

Cost

We estimate costs for this option to be approximately \$285,000.

Statistical Methods for Analyzing the Data

1. Propensity Score Matching

Propensity score matching methods will be used to generate an appropriate comparison group for the quasi-experimental evaluation design. This technique generates weights to be applied to the comparison group so that it more closely resembles the treatment group on observable variables. The match will be conducted by producing a predicted probability of being in the treatment group using a logit model of treatment group status as a function of earnings history, employment history, geographic location, industry, and other matching variables.

Ideally, matching on these observable characteristics should help to control for other unobservable characteristics as well. Certain apprentice characteristics may be correlated with the earnings of apprentices, although these characteristics are not measured in the LEHD data. While wages for direct care workers do not vary greatly (Khatutsky, Wiener, Anderson et al., 2011), the number of hours worked do, so some of the variation in earnings may reflect the unmeasured characteristics of workers selected to become apprentices.

Once a predicted probability of being in the treatment group is produced for all members of the comparison group, a variety of matching strategies can be used that predict probability to weight the comparison group. These include the nearest neighbor method, the odds ratio method, and the Kernel density method. The nearest neighbor

method pairs each treatment case with the comparison case that has the closest propensity score to it, generating a one-to-one match between the treatment and comparison group. The odds ratio method and the Kernel density method generate a weight for all comparison cases using the propensity score. Comparison cases with high propensity scores are given high weights, and those with low propensity scores are given low weights. Mueser, Troske, and Gorislavsky (2007) find that impact estimates for job training programs are not especially sensitive to the choice of matching method. In order to confirm the robustness of any evaluation of the LTC RAP initiative, multiple matching methods should be used. After implementing these matching strategies, Rubin (2001) suggests several “balancing tests” to confirm the strength of the match. The balancing tests are various versions of a difference of means test on the matching variables. A strong match should reduce statistically important differences between the treatment and comparison groups on the matching variables.

2. Impact Estimation

Once the propensity score matching method produces a viable comparison group, several estimation strategies can be used to produce an estimate of the impact of the LTC RAP program, including a difference of means test of post-registration earnings with and without regression adjustment, and a difference-in-differences test of earnings with and without regression adjustment. Mueser, Troske, and Gorislavsky (2007) find that the difference-in-differences estimator is more faithful to random assignment results, although multiple approaches should be attempted and compared. The estimated differences can be examined by the level of the propensity score; thus, one can observe changes in earnings for those most likely selected for the program as compared with changes in earnings for those least likely to be selected. Sample size constraints may limit the principal subgroup analysis to the CNA and DSS occupational categories, although other characteristics of workers (e.g., age or race) may also be of interest.

3. Alternative Versions of the Administrative Data Design

If selection bias is considered to be a major obstacle to evaluation of the LTC RAP, other alternative non-experimental strategies can be considered using LEHD data administrative data. One alternative strategy is a regression discontinuity design, which uses pre-determined cut-offs in the assignment of treatment to identify the impact of the treatment. For example, all CNAs at Agape Senior are ranked, and apprentices are chosen from among the top 20% of employees. Since there is a sharp cut-off in the assignment of treatment, cases immediately above and immediately below the cut-off are expected to be very similar on all of their characteristics, except for their admission to apprenticeship, generating a type of natural experiment. Regression discontinuity designs might be appropriate for LTC RAPs which use some sort of test or evaluation to assign employees to the apprenticeship. Only a minimal difference is expected in the performance of employees in the 79th percentile compared to the 80th percentile, but there is a large difference in likelihood of becoming an apprentice. The change in the outcome variable at this point of discontinuity provides a reasonable estimate of the

impact of the treatment. Although not common among LTC RAPs, Agape Senior is probably not the only LTC RAP program that uses an objective employee performance measure to decide who will participate (or at least who will be offered the opportunity to participate) in the apprenticeship program. If enough programs use this approach, it may be possible to use this analytic approach. Sample sizes and the frequency of using this approach of selecting apprentices may, however, limit the feasibility of this approach, especially for a national evaluation.

6.2. Survey Option for Apprentices

Survey data collection and analysis is a potential evaluation option to address research questions regarding the apprenticeship experience on topics not available in secondary data. For the LTC RAP evaluation, a survey of apprentices would provide systematic quantitative information on the experiences of apprentices with the LTC RAP compared to direct care workers not in the LTC RAP.

Research Questions

A survey of direct care workers who have participated in the LTC RAP program, including those who did not complete the program and a comparison group would address a range of research questions that cannot be answered directly by employers, sponsors, or partnering organizations or through administrative data, such as the LEHD. Potential research questions center on how the LTC RAP affects:

- Job satisfaction
- Intent to leave the job and the long-term care field
- Participation in welfare programs (SNAP, TANF, Medicaid, etc.)
- Relationship with supervisor, other staff members, and clients
- Confidence in caregiving abilities
- Knowledge and skills of caring for people with disabilities
- Time and financial investment on the part of apprentices to participate
- Opinions about the LTC RAP
- Opinions about other training for direct care work
- Future career plans

These research outcomes for the survey are generally more difficult to measure than outcomes like annual earnings, wages, and job tenure. Ideally, the outcome measures should be tested to ensure that they are valid and reliable measures and that there is variation in the responses. For example, commonly used measures of job satisfaction typically report very high levels of respondents who are “extremely” or “very satisfied” with their job (Bishop et al., 2009). Thus, it might be difficult to measure the impact of the LTC RAP on this dimension.

Brief Overview of Design

The suggested design would involve a survey of apprentices in facilities/agencies operating apprenticeship programs and a matched sample of non-apprentices at a single point in time. Thus, this survey would be a cross-sectional design; it would be able to find association between variables, but it would not be claim that the LTC RAP caused the differences because it does not measure changes over time. All direct care workers who started the apprenticeship program and still working for the facility/agency employer which administered the LTC RAP would be included. Non-apprentices would be employees of either branches within the same organization that are not implementing the apprenticeship program or employees of wholly different long-term care provider organizations not implementing apprenticeship programs.

A telephone survey is recommended to obtain data from apprentices. Given the number of sites and the small number of apprentices at most sites, an in-person survey would be prohibitively expensive. In addition, direct care workers generally have low education and literacy skills, and may also have cultural differences that make a mail survey problematic. Workers may have difficulty reading and interpreting the questions. In addition, similar surveys of CNAs (National Nursing Assistant Survey) and HHAs (National Home Health Aide Survey) have been successfully conducted by telephone.

The survey will be administered as a computer assisted telephone interview (CATI), which will ensure standardized question administration and will reduce data entry costs. Also to minimize costs, the survey would only be conducted in English and Spanish, but not other languages. The survey would be conducted over a 4-month period. Contact information, such as telephone numbers and addresses, for apprentices and the comparison group workers will be obtained from employers. As a practical matter, obtaining contact information for apprentices who have left the employment of the provider that trained them will be difficult if not impossible and will not be attempted. The survey administrator would vary the days and times of contact attempts in order to maximize the possibility of reaching sample members to schedule the full interview.

Given the similarity in the goals of the apprenticeship programs across the four occupations of the LTC RAP, and the relative few numbers of apprentices in some occupations such as HHA and HSSs, a single survey across all occupations is recommended. Even with the entire universe of apprentices, the number of completed surveys for HHAs and health care support specialists is too small to analyze separately. To control for differences across occupations, the four main LTC RAP occupations will be entered as control variables in the multivariate analyses. Subgroup analyses of CNAs, the largest occupation, and DSSs, the second largest occupation, will be possible if there is a large enough number of respondents.

Given the relatively small number of employers/sponsors and of apprentices, the sample design should include all current and past apprenticeship sponsors/employers and all apprentices currently employed by these employers/sponsors, including those who have already completed their apprenticeships and those who did not complete the

apprenticeship. The evaluator will need to identify through secondary data (such as RAPIDS) or directly through employers those apprentices that are still working for them. The comparison group of workers who have not participated in LTC RAP will be drawn from the same or, more likely, other organizations providing similar types of services.

The sample should result in approximately the same number of completed surveys for apprentices and for comparison group members. To achieve this result, the evaluator will likely need to oversample the comparison group because the response rate may be lower because of their lack of knowledge and interest in an evaluation of the LTC RAP. For prior surveys of CNAs and HHAs, the ASPE/NCHS achieved roughly 75% response rates for facilities/agencies and a 75% response rate for workers, giving a 50% overall response rate.

Conservatively, using RAPIDS data, slightly lower response rates of about 67% for the 80 current employers with approximately 1,500 apprentices in training currently would yield about 1,000 completed surveys for apprentices. Consistent with the 2004 National Nursing Assistant Survey which provided a monetary incentive to workers to encourage participation, this survey will provide a \$35 incentive payment. We do not anticipate payment of incentives to employers to provide the contact information.

The sample would include a comparison group of workers from providers not sponsoring apprenticeships or from non-apprenticeship-sponsoring branches of parent organizations who have apprenticeships in some, but not all branches. However, only a few sponsoring employers have multiple branches to make selection of comparison group members possible. Therefore, most, if not all of the comparison group would need to be drawn from non-apprenticeship-sponsoring organizations, which would have to be recruited to the study.

To provide a close comparison to apprentices and their sponsoring employers, comparison organizations ideally would be in same geographic area and have comparable size, ownership status, payer mix, and other important characteristics. This data is routinely collected for nursing homes and home health agencies by CMS, but is not available at the national level for other types of providers participating in LTC RAP. For other providers, many long-term care employers are members of state and national associations, and comparison group employers for these providers could be identified through their membership rosters. In addition, if ASPE and NCHS grant permission to use it, RTI International developed a sample frame of residential care facilities for use in the 2010 National Survey of Residential Care Facilities (Wiener et al., 2010). NCHS recently awarded a contract to RTI International to update the sample frame for residential care facilities in 2012.

Motivating non-apprenticeship-sponsoring facilities to participate will be difficult because of lack of interest or knowledge about LTC RAP or the perceived cost of participating. Moreover, facilities may not believe it is in their best interest to have outsiders asking their workers about subjects such as job satisfaction, relationships with supervisors, and wages and benefits. Employers may also be reluctant to release

personal contact information or Social Security numbers of workers without their explicit permission, even if the employers are supportive of the survey. Letters of support from provider associations and high-ranking HHS and DOL officials may help with recruitment.

Comparison group members need to closely resemble apprentices in selected characteristics. Therefore, comparison group direct care workers ideally should be prospectively matched with apprentices, potentially using employment history/earnings, age, gender, race, education or similar factors, but doing so would be difficult. Alternatively, the evaluator could control for such matching retrospectively through statistical adjustments if sufficient data were collected from both apprentices and non-apprentices. Potential response bias may still occur if important variables are not collected during the survey. For example, apprentices and non-apprentices may vary on unobserved characteristics (e.g., altruism or motivation) for which data is not collected or successfully measured.

Estimated Statistical Power

Preliminary calculations of statistical power needed to detect differences in outcomes such as satisfaction or intent to leave suggest that the apprentice group would need to have 1,000 respondents and 600 comparison group workers for a total sample of 1,600 respondents. These numbers for completed respondents are of this magnitude to allow for sufficient power for subgroup analyses of CNAs and DSSs, the two largest occupation groups. Respondent group sizes in excess of these numbers to provide enough statistical power for subgroup analyses of HHA and HSSs.

Measures such as job satisfaction and intent to leave one's job have relatively little statistical variation (Bishop et al., 2009), therefore relatively large numbers of apprentice and comparison group members are needed to detect statistically significant differences as small as 5 percentage points at probability of less than 0.05 ($p < 0.05$). For a binary outcome variable in logit analysis such as Satisfied/Not Satisfied defined on a 100 percentage point scale, one could detect a difference as small as 1.25 percentage points using a mean of 82 percentage points and a standard deviation of 10 percentage points, assuming 1,000 apprentices and 600 comparison group members. We believe a sample of this size would provide sufficient power for assessing impact.

Sample Frame Construction for Programs and Apprentices

To identify sample frame members, lead letters from important HHS and DOL officials and letters of support from the relevant provider associations would be prepared and sent to prospective employers. These letters would provide assurances that the privacy of participating employers and employees will be protected. Senior staff from the evaluator would contact employer sponsors by phone to introduce themselves and address any remaining questions and solicit commitment in participation. Once agreeing to participate, employers would provide contact information (e.g., names, telephone numbers, and addresses) of their currently employed employees who ever

participated in the LTC RAP and non-apprentice workers from other branches if possible. Similar information would be obtained from the comparison group facilities. Sample members would be sent a pre-notification letter 1 week before interviewing is scheduled to give advance notice to sample members that they have been chosen for a survey, establish survey legitimacy, and provide information about the survey.

Survey organizations often experience problems obtaining valid telephone numbers for potential respondents. Lower income people such as long-term care workers may not have listed landlines, and it is not likely cell phone numbers could be attained independently of the employers. Employers may or may not be willing to share home or cell phone numbers of workers, and it would not be appropriate to survey workers while on the job because of potential fear of retaliation by management if they criticize the facility/agency or their supervisors. Thus, the proportion of apprentices who are successfully contacted may be lower than anticipated.

Domains on Which Information Will Be Gathered

The survey will collect information on the outcomes of interest (e.g., satisfaction, intent to leave, new knowledge and skills attained) and also on an array of other domains which will be used in analyses to statistically control for factors not related to the effect of apprenticeship. These domains include:

- Worker background (e.g., demographics, socioeconomic status, family relationships, residence status).
- Personality inventory to assess fit with caregiver role.
- Employment history (e.g., number and types of previous jobs, relative prior pay and availability of benefits, previous training, life/employment skills).
- Availability and uptake of fringe benefits offered.
- Organizational culture (e.g., control over work, relationship with peers and supervisors, opportunity to work in teams, and other characteristics thought to affect satisfaction, intent to leave, and confidence in new knowledge and skills).
- Training before the apprenticeship (e.g., hours and source of basic training, whether had mentor previously).
- Views about apprenticeship (e.g., motivation for participation, what they learn, best things/worst thing, non-paid time invested, out-of-pocket costs, and views of mentorship, OJT, and related training instruction).

Questionnaire Development

The evaluator would identify the specific domains that will be included in the questionnaire and, would identify potential questions and issues related to data collection. After obtaining feedback from ASPE and DOL, the evaluator would begin to develop the actual draft questionnaire. The evaluator would prepare an Office of Management and Budget (OMB) clearance package to include the essential supporting statement sections (e.g., justification, effort to identify duplication, methods to minimize burden, cost and response burden estimates, publication plans, and statistical methodology) and relevant information on research questions and survey protocol. The final OMB clearance package will include the final questionnaire.

Data Collection Process

The survey would be conducted using a CATI system and last approximately 30 minutes. Once an interviewer makes an initial contact with a potential respondent, the interviewer will schedule a time to administer the survey. At that time, the interviewer would administer the introduction, which would include obtaining informed consent from the respondent and assures of privacy of responses. The interviewer will then administer the survey, following the script that the CATI program displays on the computer screen. The CATI system conducts edit checks for appropriate response values and correct use of skip patterns to increase data accuracy during the interview. As data are collected daily, project staff would review responses and generate frequencies and means of key variables to ensure that data look as expected and no unusual response patterns are observed. Similarly, project staff would monitor response rates daily by apprenticeship and comparison group and for the sample overall. Should response rates be lower than expected, staff would implement corrective measures, such as varying the number of call attempts or call schedule or developing more targeted scripts to address refusals or questions that sample members may have.

At the conclusion of the data collection period, the data would be cleaned (e.g., provide standardized codes for “yes”, “no”, “refusal” and “don’t know” responses) and a dataset would be created for analysis. As part of the creation of the final dataset, programmers would prepare an accompanying codebook containing questions and responses, as well as key data collection variables such as date of interview and final disposition code for any non-interviews.

Time Frame to Collect and Analyze Data

We anticipate that the entire survey option would take approximately 2.5 years to complete. The activities would include questionnaire and sample frame design (6 months), preparation of OMB package and clearance (8 months), data collection (6 months), data cleaning (2 months), and analysis and reporting (8 months).

Ballpark Cost

We estimate the total costs for conducting the survey are approximately \$450,000, which includes questionnaire and sample frame design, translation of the survey into Spanish, preparation of OMB package and clearance, data collection in English and Spanish, data cleaning, and analysis and reporting. Costs for the actual data collection would be approximately \$335,000, which include costs for programming the 30-minute, closed-item 75-question survey into the CATI system, interviewer training, developing and mailing all pre-notification letters, delivering an English and Spanish-language CATI survey over a 4-month period, multiple call attempts for approximately 10 days, a survey case management system to schedule and track calling attempts and survey status, a \$35 incentive for survey completion, cleaning the data and preparing a SAS dataset with survey frequencies and documentation of all coded items.

Main Statistical Methods for Analyzing Data

The data would be analyzed using both descriptive and multivariate regression techniques. Means for all analysis variables would be prepared for all respondents and for apprentice versus comparison group members. Descriptive analyses using comparisons of means (e.g., age) and proportions (e.g., gender) and cross tabulations of outcome measures (e.g., satisfaction, intent to leave) with selected characteristics of interest (e.g., employer profit status, worker job tenure) would be calculated. Descriptive analyses without testing for statistically significant differences could be calculated on apprentices with varying characteristics, but small sample sizes for given characteristics (e.g., those with any specialty training, various occupations, and source of related training instruction) would prevent much statistical significance testing for such differences on outcomes.

Multivariate regression would be used to analyze the effects of participating in apprenticeship on outcome measures representing the key research questions. The outcome measures are typically multilevel outcomes (e.g., extremely satisfied, somewhat satisfied, somewhat dissatisfied, extremely dissatisfied) that would be analyzed using multinomial logit, or multiple levels could be combined into only two levels and analyzed using logit. The principal independent policy variable would be a yes/no indicator of any participation in apprenticeship. The basic empirical model analyzed that would control for apprenticeship and other domains that are hypothesized to affect the outcome of interest would be:

Outcome = yes/no indicator of apprenticeship, demographic and socioeconomic status, family relationships, residence status, personality type, employment history, employer benefits, organizational culture, pre-apprenticeship training + error

There potentially may be enough CNAs and DSSs in the data to estimate regression analyses on that subgroup of apprentices, but there would not be enough apprentices in the remaining occupations to perform similar analyses. However, it is not likely that there would be sufficient statistical power to test for statistically significant

differences in apprenticeship characteristics (e.g., specialty versus only advanced competencies) among the subgroup of apprentices in regressions because few apprentices with such characteristics are likely to be represented in the data. Because we anticipate that the universe of employer/sponsors and the universe of their currently employed apprentices would be used to construct the sample frame, descriptive and multivariate analyses would not have to control for the effects of the sample design.

6.3. Focus Groups

Primary data collection using focus groups may provide a low-cost research design option that would obtain information to inform policy. Focus groups would provide a means for apprentices to voice opinions on a range of topics about how the apprenticeship program works from their perspective, its strengths and weaknesses, and how it could be improved. Focus groups of employer sponsors could provide information on why they chose an apprenticeship program and important program elements, and what they perceive to be the benefits to employers and apprentices. In both cases, focus groups allow for a more extensive, open-ended data gathering than is possible in a survey. The focus groups would also allow gathering detailed information about the LTC RAP professions for which there are too few sites to conduct statistical analyses. Because the data is qualitative and does not include a large enough sample to conduct statistical analyses, it cannot answer such quantitative questions as whether LTC RAP increases earnings or reduces turnover.

Research Questions

The research questions to be addressed by the focus groups center on the motivation for entering/starting the LTC RAP and the views of direct care workers and employers about the operation of the program.

1. Apprentices

Potential research questions to be addressed by focus groups of apprentices include:

- How and to what extent does apprenticeship affect caregiving abilities? What types of new knowledge and skills were attained? How and how well does the apprenticeship teach hard and soft skills and problem-solving skills?
- How does the apprenticeship affect working with supervisors and working in teams? How do the apprentices relate to comparable workers who do not participate in the LTC RAP?
- How does apprenticeship differ from other training previously received?
- What personal funds and time did apprentices use in order to participate?

- Were they able to complete the apprenticeship and why or why not?
- What did they like about the apprenticeship experience? What did they dislike?
- What changes would they suggest making in the apprenticeship program?
- Can LTC RAP play a major role in solving long-term care workforce problems, and if so, under what circumstances?

2. *Employers*

Research questions to be addressed by focus groups of employer sponsors would include:

- Why did the organization choose an apprenticeship program to train and develop direct care workers?
- What relationships with outside organizations were beneficial in developing your program?
- What financial, training and other resources would help the organization to best operate the program?
- What is the value added of apprenticeship over traditional training?
- What criteria does the organization use to select workers into the program?
- What do employers like about the LTC RAP? What did they dislike?
- What changes would employers suggest making in the apprenticeship program?
- Can LTC RAP play a major role in solving long-term care workforce problems, and if so, under what circumstances?

Brief Overview of Research Design

1. *Apprentice Focus Groups*

This evaluation design would conduct multiple apprentice and employer/sponsor focus groups. First, eight apprentice focus groups in eight sites would be conducted. The evaluator would recruit apprentices from selected employer sponsors who would provide the names and contact information for all of their employed apprentices. Employers would not be allowed to select the participants. Two focus groups would be conducted for each of the four LTC RAP occupations. Conducting more than one focus group for each occupation and conducting multiple focus groups across employers

minimizes the possibility that results are purely idiosyncratic or reflect the experience of just a few participants who are the most verbal in an individual group. Final site selection of employer sponsors would occur after discussions with ASPE and DOL. Possible selection criteria for identifying employer sponsors from whom apprentices would be selected would be total number of apprentices employed, occupation type, ownership type and chain status.

Each focus group will consist of approximately 8-9 worker participants, for a total of approximately 70 participants. Budget assumptions are based on recruiting English speaking participants in order to avoid the added costs associated with translating the focus group protocol into the additional language and translating the focus group discussion into English. Even though 70 participants would be included, OMB clearance would probably not be needed because each focus group would contain a maximum of nine individuals, and question scripts would necessarily differ across occupations and by employer so that no set of focus group questions would be the same.

The evaluator would work with each employer to help arrange for local logistics for conducting each focus group. In order to encourage workers to be candid, the focus group will be held at a location away from the employer's work site, but convenient for the apprentices. A small token of appreciation would be provided to the employer for their efforts in providing the names and contact information of the apprentices and suggesting focus group locations. A \$75 payment would be provided to focus group members as an incentive to participate.

No comparison group of non-apprentices would be involved. With such small numbers of participants, no statistical analysis would be performed.

2. Employer Focus Groups

Because of the difficulty and expense of bringing employer sponsors around the country together for a focus group, only two focus groups of LTC RAP employer sponsors is proposed. For convenience, this focus group would occur as a side meeting of long-term care national association annual meetings. For example, the American Health Care Association and LeadingAge are two of the largest national associations of nursing homes and residential care facilities. Currently, many LTC RAP employer sponsors are members of these organizations. Both organizations have annual meetings at which a focus group potentially could be conducted.

In addition to invitations to employer sponsors from the evaluator, letters from officials of the national associations and high ranking HHS and DOL officials encouraging organizational members to participate in the focus groups would be included to promote the importance and legitimacy of the endeavor. The evaluator would select up to nine employer sponsors from each association who would attend separate focus groups. Once contacted, the evaluator would work to identify the best day and time across all selected employer sponsor personnel attending their respective national association meetings. Employer sponsors would not incur additional costs for

attending the focus group session beyond their time to prepare and attend. Therefore, only some nominal incentive payment to participants is anticipated.

Sample Frame Construction

For the apprentice focus groups, the entire universe of employers and currently employed apprentices, all identified from the most recent RAPIDS data extract, would serve as the sample frame. Employers would be stratified by their LTC RAP occupations and number of employed apprentices, and on selected apprenticeship program characteristics thought to be instrumental in conducting apprenticeships such as ownership status. Once employers are identified, all currently employed past and current apprentices would be eligible for recruitment. For the employer sponsor focus groups, the evaluator would work with association management to identify a convenience sample of LTC RAPs in RAPIDS data who are members of the respective associations.

Domains on Which Information Will Be Gathered

To guide discussion, the evaluator would develop a focus group protocol and structured discussion guides, including specific questions with suggested probes. The protocol would include discussion of approximately 7-8 topics and last about 2 hours. Questions for apprentices would be prepared recognizing the limited education of most apprentices. Topic domains would include those addressed in the research questions previously listed.

Data Collection Process

As part of the enlistment process, participants will receive letters and phone calls asking them to participate. Provisions for informed consent and confidentiality of responses would be provided as part of the process. Any special needs of enlisted participants would be identified. Details of the logistics (e.g., date, time, and location) of the focus group would be provided to participants.

The day before the focus group, two staff members from the evaluator would travel to each site to preview the site, make any final arrangements, and setup. Upon arrival, focus group participants would be greeted, registered, introduced to other participants, and invited into the room. Light refreshments and food would be provided. The focus group moderator would provide a brief overview of the process, establish rapport with participants, and conduct the focus group. The second staff person would take notes on a laptop and provide assistance for issues that arise during the group so the moderator could continue with group moderation. The focus group will be recorded; only apprentices who agree to have their comments recorded will be included in the focus group. Generally, it is assumed that given the physically demanding job of apprentices, none would need extensive accommodations during the meeting. For the focus group for employer sponsors, any special accommodations most likely would be provided as part of their association meeting.

Timeframe to Collect and Analyze Data

Approximately 12-14 months would be needed to identify employers and sites, develop the necessary materials, recruit apprentices and leadership staff from employer sponsors, conduct the focus groups, analyze the data, and develop a research report of findings. If OMB clearance is required, then an additional 6 months would be required to complete the project.

Ballpark Estimated Cost

The entire focus group evaluation option would cost approximately \$150,000, which includes labor and travel costs to develop the protocol, conduct the focus groups, analyze the focus group data, and write a single summary report of findings. The subset of costs for actual conduct of each focus group would be approximately \$10,000 for recruitment from a list, logistics, facilitation and written transcripts.

Main Methods for Analyzing Data

The transcripts of the focus groups would be analyzed for themes either by hand or through a qualitative analysis software package such as QSR NVivo. A single thematic report would be developed along the lines of the topic domains contained in the discussion guide. One section of the report would discuss apprentice findings, and a second section would discuss employer sponsor findings.

6.4. Cost-Benefit Analysis

Employers in long-term care are highly constrained in what they pay their workers, face government training requirements even if they do not offer apprenticeships, and often pay low-wage rates even to highly qualified workers. As in any service industry, productivity in long-term care is hard to measure. Indeed, in some cases, increased “efficiency” is an indicator of poorer quality (e.g., reduced feeding time for severely disabled nursing home residents). For any voluntary program, like the LTC RAP, that provides more extensive training than that required by federal and state law and regulation, long-term care providers must believe that the additional training is worthwhile for the organization. In other words, the benefits must exceed the costs. This component of the evaluation will quantify those factors for employers participating in the LTC RAP.

The primary resource costs are the salaries of instructors, mentors, office staff, and others for the periods they deliver the training and the dollar costs of materials, rent, and other material inputs. Another cost is the lost productivity of participants while they are taking part in the training, a cost which is minimized in apprenticeship programs. The primary benefits to the employers is reduced turnover and improved productivity/quality of care provided by the trainees, usually measured in evaluations of

job training programs as the present value of the increases in earnings over time. Improved quality may reduce accidents and increase the number of consumers who use the service.

Research Questions

- Given these considerations, the key questions in the cost-benefit analysis are:
- What is the impact of LTC RAPs on employer costs?
- What is the value of benefits of long-term care apprenticeships to employers?
- What is the ratio of benefits to costs for employers?
- What costs do apprentices incur by participating in a long-term apprenticeship?
- What is the impact of participating in an apprenticeship on subsequent employment and earnings?
- What is the ratio of benefits to costs to workers of participating in LTC RAPs?

Overview of Design

This evaluation option will conduct a cost-benefit analysis from the perspective of LTC RAP employer sponsors and from the perspective of apprentices. Since little direct public money is spent on LTC RAP, the analysis will not be from the societal government perspective. Data on the employers will be collected through a web, mail and telephone follow-up survey and structured discussions with the management of LTC RAP employer sponsors. The option would not include a comparison group. Data generated from the LEHD data analysis will be used to conduct the cost-benefit analysis for apprentices.

1. Employers

Through in-depth interviews and surveys, the evaluator can work with employers to determine cost elements, cost savings, and other benefits to LTC RAP sponsors. The survey questions would be explained carefully to the employer and the evaluator would provide technical support to clarify any questions the employer might have. The evaluation option will collect data on the costs and benefits of the LTC RAP from 50 employer/sponsors. Forty of the employers/sponsors will complete a web survey. In addition, ten of the largest LTC RAP employer sponsors would receive a more detailed telephone survey to gain a deeper understanding of their costs and benefits. The LEHD administrative record analysis will provide estimates of the extent to which LTC RAP apprentices leave their employers at rates similar to or different from other workers with similar wages in the firm. This information will strengthen estimates of the potential benefits employers derive from reduced turnover.

LTC RAP employers will be asked to identify the time costs of trainers, the costs of related instruction, the wage costs of the apprentices while in class training, and the cost of replacing apprentices while they are in class training. They will be asked to specify whether the apprentices are paid for the time they invest in related instruction. In some cases, other organizations, government agencies, and foundations reimburse some employer costs associated with apprenticeships. Nonetheless, these costs are true costs regardless of who pays for the use of resources and should be included.

On the benefit side, employers will be asked to report turnover for various classes of workers. LTC RAP employers will be asked to supply data on turnover for those who took apprenticeships and those who did not, and turnover rates before and after the adoption of apprenticeship training, if available. If apprenticeships do lead to fewer quits or discharges, the monetary benefit of reductions in turnover will be calculated. To do so, employers will be asked to quantify recruitment and initial training costs for regular workers not in an apprenticeship program. In addition, the apprenticeship program might lower the likelihood of errors or accidents. Employers of long-term care workers would be asked to make informed guesses based on their experience about the potential cost savings from reductions in errors or accidents. These estimates will have a high-level of measurement error and will be used only to develop gross order of magnitude estimates.

2. Apprentices

In order to conduct the cost-benefit analysis from the perspective of the apprentice, data from the LEHD will be used. We propose that the evaluator follow the procedure of Hollenbeck (2011) in estimating foregone earnings and post-apprenticeship net earnings. In the case of long-term care, wages have little variation, resulting in modest benefits from the standpoint of wage gains. However, the advantages of apprenticeship for the workers may be significant in terms of retaining employment and the stability of employment within the firm and within the industry and in being assigned to work more hours as a more highly valued staff person. Thus, earnings of workers may rise as a result of apprenticeship programs even if hourly wage rates do not increase very much. The empirical evidence drawn from the LEHD-based impact study will yield estimates of earnings differentials between apprentices and matched comparison groups in each calendar quarter from the point of entry into the apprenticeship program through the latest post-program period.

The analysis will analyze earnings during the apprenticeship period to determine whether apprentices have foregone earnings by participating in the program. Based on the site visits, apprentices will not incur any foregone earnings within the long-term care firm for which they work, but they might forgo earnings that might have been achieved by leaving for a new job at another provider or leaving the long-term care field altogether. In some cases, apprentices incur out-of-pocket costs and they do forgo some leisure time to participate in classes and to study. Next, the analysis will calculate estimates of earnings gains achieved after completing the program. Projecting gains far

beyond the period for which data are available involves uncertainties, but evaluators can provide sensitivity analyses to determine how the size of net benefits for apprentices varies with what one assumes about whether any observed gains in earnings erode over time after the last follow-up period. Once each quarter's earnings impact is determined, the evaluators will calculate the present value of earnings gains.

Domains on Which Information Will Be Gathered and Questionnaire Development

The employer survey will gather information on the costs and benefits to employers. The primary costs to employers are:

- Salary costs per hour of the mentor/teacher times the number of lost production hours
- Wages of the apprentice lost to production
- Employer spending on classroom instruction
- Management salary and other costs for administering the program
- Miscellaneous administrative costs (e.g., reporting to DOL, record keeping, award certificates, etc.)

Weighed against these costs are the benefits to the employer. The primary benefits are:

- The value of production generated by apprentices during the apprenticeship period.
- The ability for employers to obtain post-program benefits because the productivity of workers completing their apprenticeships exceeds their wage. This result can occur because of imperfect information (the training firm knows the worker better than other firms) or other factors (Wolter and Ryan, 2011).
- Savings in recruitment and other training costs associated with lower turnover.
- Increases in the quality of goods and/or services resulting from a more highly trained workforce; may include quality improvements that reduce risks of high cost mistakes.
- Savings in worker compensation and insurance premiums.
- Improved reputation for quality as high-level training will be perceived by consumers as a proxy for higher quality services.

Evaluations of some job training programs and particularly apprenticeship programs ask employers to examine the effect of the training on productivity. ***Exhibit 9*** provides an example of a typical question for assessing productivity gains. This question and other questions form part of the existing self-assessment tools used in

other countries by which employers can examine the benefits and costs of their apprenticeship program.

EXHIBIT 9. Example of Question Used to Assess Productivity Gains				
How does the productivity of apprentices compare to the productivity of a worker who has achieved mastery in the occupation?				
<i>Please take time to answer this question and consider the average values of the past years.</i>				
Handling productive and challenging issues on the job is considered to lead to a better development of professional competence. At the beginning of their apprenticeship, trainees can accomplish only some tasks normally undertaken by master employees. Please indicate the productivity of apprentices in percentage terms compared to the productivity of skilled workers.				
Example: 50% is equivalent to “apprentice is half as productive as a master employee”				
100% is equivalent to “apprentice is as productive as a master employee in this occupation”				
	At the end of each semi-annual interval			
Apprentice	1	2	3	4
Productivity				
In percent of a master employee	___%	___%	___%	___%
SOURCE: Adapted from the Quality Returns and Costs (Form 6 Cost-Benefit Analysis) an on-line tool for company self-assessment developed by Ursel Hauschildt and Felix Rauner at the TVET Research Group, Institute for Technology and Education, University of Bremen.				

Time Frame to Collect and Analyze Data

The main tasks are: (1) developing the survey instrument, creating an on-line version and pretesting the survey; (2) obtaining OMB clearance; (3) conducting data collecting and providing technical assistance to employers as they complete the survey with 40 employers/sponsors; (4) 1 hour in-depth telephone interviews with ten of the employers/sponsors; (5) analysis of the data; and, (6) write up of the cost-benefit analysis. It is estimated that, including 6 months for OMB clearance, that the cost-benefit analysis will take 14 months.

Cost Estimate

The estimated cost for conducting the cost-benefit analysis is \$100,000.

Main Statistical Methods for Analyzing Data

The standard method for estimating the value of production is to ask employers about the productivity of apprentices relative to the productivity of workers who have fully mastered the occupation. On the assumption that master workers contribute to production what they earn as wages, the production benefits equal:

$$VP_a = W_q * (P_a/P_q) * H_a,$$

where VP_a is the value of the added production generated by the apprentice, W_q is the wage rate of master workers, P_a/P_q is the productivity of the apprentice relative to the productivity of master workers, and H_a is the hours the apprentice devotes to production.

Additional Analyses

Based on the value of production and other benefits (e.g., reduced hiring and training costs) relative to costs, it is possible to calculate the present value of benefits and costs during the apprenticeship period and during the apprenticeship and 3-4 years of post-apprenticeship for each employer. The report will then display the distribution of net benefits (benefits - costs) and the cost-benefit ratios across employers. For example, it will show what share of firms reaped net benefits of say, \$4,000 and over, \$3,000-\$3,999, \$2,000-\$2,999, ... negative 1,000-negative 1,999, less than negative \$2,000, etc.

The cost-benefit analysis can examine the relationship between the characteristics of employers and their programs and the net benefits they accrue from a LTC RAP. Other studies have documented high-levels of employer satisfaction from their participation in registered apprenticeship programs (Lerman, Eyster, and Chambers, 2009). To the extent that the results indicate high net benefits in this apprenticeship program, the findings can be used to help market the program to the long-term care industry as a whole.

7. CONCLUSION

The long-term care industry faces shortages of highly trained direct care workers. As a result, the industry struggles to improve quality of care and the lives of workers, who often receive low wages and few fringe benefits. Jobs, such as CNAs and HHAs, are often viewed as dead-end jobs, with little opportunity for career advancement. While apprenticeship has a long and successful history in other occupations and in other countries, its application to long-term care is fairly new. Understanding whether apprenticeship can address the industry's workforce shortage and increase quality of care and worker future prospects is unknown.

An important element in determining apprenticeship's future in long-term care is evaluating the effects of the LTC RAP for employers and workers. In order for policymakers, direct care workers, and employers to decide whether to promote and participate in the LTC RAP depends on the benefits and costs, both monetary and non-monetary, to both groups. Potential evaluation approaches must be considered in the context of the current status of apprenticeship in long-term care organizations, how it is implemented, and the likelihood of its future dissemination among employers.

The analysis of RAPIDS data for this project (Anderson et al., 2010) and the site visits to program sponsors (Kuehn et al., 2011) provided information to identify a range of research questions at both the apprentice and employer-levels. To address these questions, RTI International and the Urban Institute examined a broad range of potential research design options to measure the effects of apprenticeship on outcomes related to these questions. The research designs varied in terms of their ability to provide generalizable findings and in their implementation costs.

Although LTC RAP apprenticeships usually have the same goals of improving long-term care quality through a better trained workforce and thus could potentially be evaluated as a whole, the nature of LTC RAPs poses certain constraints on the ability to implement any given evaluation design. For example, the LTC RAPs are relatively small, have high apprentice turnover rate, include lengthy training periods, feature purposeful selection of better than average employees, and provide limited data collected by employer sponsors on outcomes. These program features pose considerable challenges to most research design options. In addition, the difficulty in identifying appropriate comparison groups threatens the validity of many of the research designs considered.

After careful consideration, the RTI International/Urban Institute team identified four potential research designs that together form a comprehensive approach to an evaluation of the LTC RAP. These designs include apprentice worker or employer sponsor-level analyses. They span the range of less generalizable, less expensive qualitative designs to more generalizable, more expensive multivariate analyses. As

shown in **Exhibit 10**, each has certain strengths and weaknesses. If all four components were funded, the estimated cost would be \$985,000.

EXHIBIT 10. Overview of Potential Evaluation Design Options to Evaluate the LTC RAP		
Option	Advantages	Disadvantages
<p>Analysis of LEHD, comparing all apprentices with matched sample comparison group \$285,000 27 months</p>	<ul style="list-style-type: none"> • Uses data on all apprentices, regardless of when they started and whether they completed the program • Captures duration with the firm before, during, and after apprenticeship • Addresses major issues of earnings and job tenure and continued employment in the industry • Dataset likely to include very high percentage of people ever participating in LTC RAPs • Easy access to a large supply of low-earning people working for non-apprentice long-term care providers for comparison group • No new data collection required; no OMB review required 	<ul style="list-style-type: none"> • Limited data on which to match apprentices and comparison group, leaving possibility of uncontrolled for selection bias • No data from perspective of apprentices on outcomes such as job satisfaction • No data from perspective of employers, except for duration of apprentices within the firm • Low-wage workers in comparison group will include housekeepers and dietary staff as well as direct care workers
<p>One-time cross-sectional survey of apprentices and matched comparison group \$450,000 30 months</p>	<ul style="list-style-type: none"> • Addresses more subjective outcomes, such as job satisfaction and relationship with supervisor • Provides more detailed data on apprentices • Possible to more completely control for selection bias 	<ul style="list-style-type: none"> • As cross-sectional design, only able to analyze “association” rather than causation • Comparison group facilities/agencies may be reluctant to provide contact information about apprentices • Correction for selection bias can only made after initial contact since providers unlikely or unable to provide detailed information on workers, raising costs • Only able to include apprentices who have stayed with employer that trained them; apprentices that left employer or field lost to analysis • Less consensus on measurement of “softer” outcomes • More expensive than other options

EXHIBIT 10 (continued)		
Option	Advantages	Disadvantages
Focus groups of apprentices and of employers \$150,000 14 months	<ul style="list-style-type: none"> • Low-cost option • Provides information on views of apprentices • Can provide detailed suggestions from participants for improving LTC RAPs 	<ul style="list-style-type: none"> • Qualitative data cannot be used to determine effectiveness of intervention • Representativeness of views expressed cannot be directly assessed • Views by apprentices and providers provided cannot be easily summarized or quantified • Comparisons cannot be made to workers who did not participate in LTC RAP
Cost-benefit analysis \$100,000 14 months	<ul style="list-style-type: none"> • Attempts to measure whether benefits to employer exceeds the costs, which is key for establishing business case for program • Measures changes in turnover related to LTC RAPs • Consistent with approaches used in other studies of apprenticeship costs and benefits • Low-cost data collection 	<ul style="list-style-type: none"> • Measurement of relative productivity of apprentices is not straight forward • Employer estimates may be biased as some try to justify their investments

The first design option would use the LEHD administrative database to provide findings that could be useful for making the business case to employers, if the findings for employers and workers were positive. The LEHD is a Census Bureau database that includes state-level Unemployment Insurance administrative information on employment and earnings merged with certain other data. This design option would assess the effect of LTC RAPs on increased apprentice earnings and job tenure, and on the worker turnover rate at the employer level. Since an administrative dataset would be used, no additional data collection would be required and few employers or workers should be missing from the dataset. As an administrative dataset designed for other purposes, no data would be available about factors such as job satisfaction, relationships with supervisors, or quality of care provided. The biggest challenge for this design is using the limited variables available in the Unemployment Insurance data to construct a truly comparable comparison group. However, in other studies, prior earnings have been used effectively to proxy many personal characteristics. Although pre-program wage rates may vary little among potential apprentices and comparison group members, hours and weeks worked do vary considerably. In addition, the data can identify low-wage workers in other long-term care organizations that do not run LTC RAPs, but it cannot separate direct care workers from other low-wage workers. In particular, in residential settings, the analysis cannot differentiate between direct care workers and housekeeping and dietary staff, thus making the comparisons with apprentices somewhat imprecise, although matching on earnings should eliminate most of the non-direct care workers. Despite this limitation, this option provides potentially the most

viable design to credibly address the most important research questions facing the industry. The estimated cost of this option is \$285,000.

The second design option is a cross-sectional, one-time telephone survey of apprentices and a comparison group of non-apprentices to determine the effects of apprenticeship on job satisfaction, intent to leave one's job, relations with supervisors and other staff, and other factors that only workers can address. The survey findings would be analyzed using statistical techniques that would identify the association of apprenticeship with job satisfaction and intent to leave, but causality could not be attributed to the LTC RAP because there are no measures of change over time. While the survey could include variables that would better control for selection bias, the survey is not as well suited as the administrative data option at assessing the economic consequences of LTC RAP for apprentices. Moreover, among direct care staff that stay in their jobs, previous studies have found high rates of job satisfaction, suggesting either that existing measures are not very sensitive or that there is not that much room for improvement among workers who stay in their jobs (Bishop et al., 2009). The cost for this option is relatively high because of the expense of data collection; the estimated cost for this option is \$450,000.

The third design option would provide a much more detailed understanding of apprentice and employer opinions about how apprenticeship works. Eight focus groups would be conducted among apprentices at eight different employers, and two focus groups would be conducted among management of employer sponsors while they are attending national provider association meetings. These focus groups would provide a rich understanding of the value of apprenticeships over traditional training and how employers implement their LTC RAPs, but it could not provide any quantitative estimates of the effect of LTC RAPs. Next to the cost-benefit analysis, this option is the lowest cost option. The estimated cost of this approach is \$150,000.

The fourth evaluation design option focuses on employer-level benefits and costs of LTC RAPs. Benefits, measured as the increased productivity achieved by the LTC RAP, and a range of implementation costs would be gathered through an Internet interview process among a selected group of employers. Data from the LEHD analysis would also be used to determine benefits. Costs would include supervision costs, the time lost to regular work, and whatever curriculum development that the facility does. One challenge to this design is posed by the need for employers to accurately assess the improvement in performance and productivity due to the LTC RAP. In addition, if limited to relatively large LTC RAPs, the analysis will not include a large enough number of employers to generalize across all LTC RAPs. Still, this design would address questions related to the business case for employers, at a much lower price than the LEHD or survey design options. The estimated cost of this option is \$100,000.

In considering these alternatives, ASPE/HHS and DOL must answer two major questions. First, can the LTC RAP be a strong enough intervention to yield net benefits at the apprentice or employer/sponsor-level? Is it plausible to expect gains in increased wages, job tenure, job satisfaction, commitment to the industry, productivity and quality

of care and decreased turnover as a result of participation in the LTC RAP? In other words, can the LTC RAP approach plausibly improve outcomes for consumers, workers, employers, clients and funders for a large number of apprentices and employers? Second, can the research designs presented here or other possible designs produce findings that can withstand critical scrutiny from researchers and policymakers? In other words, will the evaluation provide methodologically defensible results that justify the cost of the evaluation?

8. REFERENCES

- American Health Care Association. (2010). Report of findings 2008: Nursing facility staff vacancy, retention and turnover survey. Washington, DC: American Health Care Association. Available at: http://www.ahcancal.org/research_data/staffing/Documents/Retention_Vacancy_Turnover_Survey2008.pdf.
- Anderson, W., Khatutsky, G., Wiener, J.M., Lerman, R., and Kuehn, D. (2010). *A Descriptive analysis of the U.S. Department of Labor's long-term care registered apprenticeship program*. Research Triangle Park, NC: RTI International and Urban Institute. Available at: <http://aspe.hhs.gov/daltcp/reports/2010/LTCappre.shtml>.
- Bishop, C., Squillace, M., Meagher, J., Anderson, W.L., and Wiener, J.M. (2009). Nursing home work practices and nursing assistants' job satisfaction. *Gerontologist*, 49(4):611-622. Available at: <http://aspe.hhs.gov/daltcp/reports/2009/NHwork.htm>.
- Blatter, M., Mu"hleemann, S., Schenker, S., and Wolter, S.C. (forthcoming 2011). Hiring costs for skilled workers and the supply of firm-provided training. *IZA Discussion Paper*.
- Hollenbeck, K. (2011). Short-term net impact estimates and rates of return. In *The Workforce Investment Act: Implementation experience and evaluation findings*. Edited by D. Besharov and P. Cottingham. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.
- Institute of Medicine. (2008). *Retooling for an aging America: Building the health care workforce*. Washington, DC: National Academies Press.
- Johnson, R.W., Toohey, D., and Wiener, J.M. (2007). *Meeting the long-term care needs of the baby boomers: How changing families will affect paid helpers and institutions*. Washington, DC: Urban Institute. Available at: http://www.urban.org/UploadedPDF/311451_Meeting_Care.pdf.
- Khatutsky, G., Wiener, J., Anderson, W., Akhmerova, V., Jessup, E.A., and Squillace, M.R. (2011). *Understanding direct care workers: A snapshot of two of America's most important jobs: Certified nursing assistants and home health aides*. Waltham, MA: RTI International. Available at: <http://aspe.hhs.gov/daltcp/reports/2011/CNAchart.htm>.
- Kuehn, D., Lerman, R., Eyster, L., Anderson, W.L., Khatutsky, G., and Wiener, J.M. (2011). *Characteristics of long-term care registered apprenticeship programs: Implications for evaluation design*. Washington, DC: Urban Institute and RTI International. Available at: <http://aspe.hhs.gov/daltcp/reports/2011/LTCRAPch.shtml>.

- Lerman, R., Eyster, L. and Chambers, K. (2009). *The benefits and challenges of registered apprenticeship: The sponsors' perspective*. Washington, DC: U.S. Department of Labor. Available at: http://www.urban.org/UploadedPDF/411907_registered_apprenticeship.pdf.
- Mansfield-Loynes, K. (2011). *Apprenticeships: What's the point?* MA Thesis. West Midlands, UK : University of Wolverhampton Business School.
- Mueser, P., Troske, K. and Gorislavsky, A. (2007). Using state administrative data to measure program performance. *The Review of Economics and Statistics*. 89(4):761-783.
- Muñhlemann, S., Schweri, J., Winkelmann, R., and Wolter, S.C. (2007a). An empirical analysis of the decision to train apprentices. *Labour: Rev. Lab. Econ. & Ind. Relations* 21(3):419-441.
- National Center for Assisted Living. (2010). *Findings of the NCAL 2009 assisted living staff vacancy, retention, and turnover survey*. Washington, DC: National Center for Assisted Living. Available at: <http://www.ahcancal.org/ncal/quality/Documents/2009NCALVacancyRetentionTurnoverSurveyReport.pdf>.
- Rubin, D. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology*. 2:169-188.
- Squillace, M., Remsburg, R., and Bercovitz, A. (2006). *An introduction to the National Nursing Assistant Survey*. Washington, DC: U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation. Available at: <http://aspe.hhs.gov/daltcp/reports/nltcssu2.htm>.
- Stone, R., and Wiener, J. (2001). *Who will care for us? Addressing the long-term care workforce crisis*. Washington, DC: Urban Institute. Available at: <http://aspe.hhs.gov/daltcp/reports/ltcwf.htm>.
- U.S. Department of Labor, Bureau of Labor Statistics. (2010). *Occupational outlook handbook, 2010-11 edition*. Home health aides and personal and home care aides. Washington, DC. Available at: <http://www.bls.gov/oco/ocos326.htm>.
- U.S. National Center for Health Statistics. (undated). *National Home and Hospice Care Survey and National Home Health Aide Survey*. Hyattsville, MD: National Center for Health Statistics. Available at: http://www.cdc.gov/nchs/data/nhhcsd/NHHCS_NHHAS_web_documentation.pdf.

Wiener, J. M., Freiman, M. P., and Brown, D. (2007). *Strategies for improving the quality of long-term care*. Washington, DC: National Commission for Quality Long-Term Care.

Wiener, J. M., Lux, L., Johnson, R. & Greene, A.M. (2010). *National survey of residential care facilities: Sample frame construction and benchmarking report*. Washington, DC: RTI International. Available at:
<http://aspe.hhs.gov/daltcp/reports/2010/sfconst.htm>.

Wiener, J. M., Squillace, M. R., Anderson, W. L., & Khatutsky, G. L. (2009). Why do they stay? Job tenure among certified nursing assistants in nursing homes. *Gerontologist*, 49(2):198-210. Available at:
<http://aspe.hhs.gov/daltcp/reports/2009/whystay.htm>.

Wolter, S. C. and Ryan, P. (2011). Apprenticeship. In *Handbook of the economics of education*. Volume 3. Edited by E. Hanushek, S. Machin, and L. Woessman. Amsterdam: Elsevier.

Design Feasibility Study to Evaluate the Long-Term Care Registered Apprenticeship Program

Reports Available

A Descriptive Analysis of the U.S. Department of Labor's Long-Term Care Registered Apprenticeship Programs

HTML

<http://aspe.hhs.gov/daltcp/reports/2010/LTCAppre.shtml>

PDF

<http://aspe.hhs.gov/daltcp/reports/2010/LTCAppre.pdf>

Characteristics of Long-Term Care Registered Apprenticeship Programs: Implications for Evaluation Design

Executive Summary

<http://aspe.hhs.gov/daltcp/reports/2011/LTCRAPches.shtml>

HTML

<http://aspe.hhs.gov/daltcp/reports/2011/LTCRAPch.shtml>

PDF

<http://aspe.hhs.gov/daltcp/reports/2011/LTCRAPch.pdf>

Evaluation Design Options for the Long-Term Care Registered Apprenticeship Program

Executive Summary

<http://aspe.hhs.gov/daltcp/reports/2011/LTCRAPedoes.shtml>

HTML

<http://aspe.hhs.gov/daltcp/reports/2011/LTCRAPedo.shtml>

PDF

<http://aspe.hhs.gov/daltcp/reports/2011/LTCRAPedo.pdf>

To obtain a printed copy of this report, send the full report title and your mailing information to:

U.S. Department of Health and Human Services
Office of Disability, Aging and Long-Term Care Policy
Room 424E, H.H. Humphrey Building
200 Independence Avenue, S.W.
Washington, D.C. 20201
FAX: 202-401-7733
Email: webmaster.DALTCP@hhs.gov

NOTE: All requests must be in writing.

RETURN TO:

Office of Disability, Aging and Long-Term Care Policy (DALTCP) Home
http://aspe.hhs.gov/_office_specific/daltcp.cfm

Assistant Secretary for Planning and Evaluation (ASPE) Home
<http://aspe.hhs.gov>

U.S. Department of Health and Human Services (HHS) Home
<http://www.hhs.gov>